

hmdr1	1	MDLEGRNGGAKKNF....FKLNKSEKDKKEKPT..VSFSMFYRNWLDKLYMVVGTAAIHHGAGLPLMMLVFGEMTDIFANAGNLEDLMSITNRSINDTGF
atpac	1	~~~~~MSETNTDAKTPAEAEKKKEQSLFFKLFSPADKFDYLLMFVGSLSGAIHVGSSMPVFFLLFGQMVNGGKNQMDL.....s
consensus	1	md e g a l s dr kkk vgv lFryadw Dk1 M lctlaaiHGs lPlmmivFgemtd fa
hmdr1	105	MN..LEEDMTRYAYXSGIGAGVLVAAAYIQVSWCLAAQROHKKIRKQFFHAINRQEI GWFDVH.DVGELNTRLTDDVSKINEVIGDKIGMFFQSMATEFTFGIVGFTTRG
hmdr1	102	SNSSLEEEEMAIYAYYYTGIGAGVLIVAYIQVSLWCLAAQROHKKIRKQFFHAINRQEI GWFDVH.DVGELNTRLTDDVSKINDGIDGDKIGMFFQSIITFLAGTIFIGFISG
atpac	77	..HQMVEHVSRYSLYFVLGLVVCFSYSAEIAQWYSGERQVAAALRKRYLEAVLKQDVQFFDTDARTGDIVSFSTDTLLVQDAISEKVGNFHYLSFTFLAGLVVGVSA
atpgp1	80	..EKMEEVLYKALYFLVVGAIIWASWAEISWMSGERQTKMIRIKYLEAALNQDIQFFDTEVTSDDVFAINTDAVMVQDAISEKLGNFHYMATFVSGFTVGVFTAV
atpgp2	73	..KOASHRVAKYSLDFVYLSVAIFLSSWLEVACWHTGERQAKMRRAYLRSLMSQDISLFDTEASTGEVISAITSIDLVDQDAISEKVGNFHYISRFIAGFAIGFTSV
consensus	111	k leemtrYayyyvglgagvlv ayiqvs W laagRQirkir kfhailrQeigwFdi tgelnltrlttdiskindgigKvGmfq vatFlagfivGfi g
hmdr3	214	WKLTLVIMAI SPILGLSAAVWAKILSAFSDKELAAAYAKAGAVAEALGAIRTVTAFGGQNKELERYQKHLENAKEIGIKKAI SANISMGIAFLIIYASALAFWYGSTLV
hmdr2	211	WKLTLVIMAI SPILGLSTAVWAKILSTFSDKELAAAYAKAGAVAEALGAIRTVTAFGGQNKELERYQKHLENAKIGIKKAI SANISMGIAFLIIYASALAFWYGSTLV
hmdr1	212	WKLTLVILAI SPVLGLSAAVWAKILSTFSDKELAAAYAKAGAVAEALGAIRTVTAFGGQNKELERYKNLENAKIRIGIKKAITANI SIGAFLIIYASALAFWYGSTLV
hmdr1	211	WKLTLVILAI SPVLGLSAAVWAKILSTFSDKELAAAYAKAGAVAEALGAIRTVTAFGGQNKELERYKNLENAKIRIGIKKAITANI SIGAFLIIYASALAFWYGSTLV
atpac	185	WKLALISVAVIPGIAFAGGLYAYTLTGITSKSRESYANAGVIAEQAIQVTVYSYVGESKALNAYSDAIQYTLKLGKAGMAKGLGLGCTYGIACMSWALVFWYAGVFI
atpgp1	188	WQLALVTLAVPLI AVIGGIHTTTL SKLSQESLSQAGNIVEQTVVQIRVMAFVGESRASQAYSSALAKIAQKLGKGTGLAKGMGLGATYFVVFCCYALLWYGGYIV
atpgp2	181	WQISLVTLISVPLIALAGGIYAFVAIGLIARVRKSYIKAGEIAEAEVIGNVRTVQAFTGEERAVRLREALENTYKGRKAGLTGKGLGSMHCVLFLSWALLVWFTSVVV
consensus	221	WkltlvillaisPiiglsaavwakils fs kel ayakAGavaEe lgairtViafgGq keleriqk le akkiGiKkaisa ismG afliiyasyAlafwygstlv
hmdr3	324	ISKEYTIGNAMTVFFSILIGAFSVGQAAPCIDAFANARGAAAVIFDIIDNNPKIDSFSERGHKPDSTIKGNLEFNDVHFSYPSRANVKILKGLNLKVQSGQTVALVGS
hmdr2	321	ISKEYTIGNAMTVFFSILIGAFSVGQAAPCIDAFANARGAAAVIFDIIDNNPKIDSFSERGHKPDSTIKGNLEFNDVHFSYPSRANVKILKGLNLKVQSGQTVALVGS
hmdr1	322	LSGEYSIGQVLT VFFSVLIGAFSVGQAAPSIDAFANARGAAVEIFKIIDNKPSIDSYSKSGHKPDNITKGNLEFSDVHFSYPSRANVKILKGLNLKVQSGQTVALVGS
hmdr1	321	LSNEYSIGEVLT VFFSILIGTFSIGHLAPNIEAFANARGAAVEIFKIIDNEPSIDSFSKYGKPDSTIMGNLEFKNVHFNYPSPRSEVQILKGLNLKVQSGQTVALVGS
atpac	295	RNGQTDGKFAFTAI FSAIVGMSLGQSFSLGAFSGKAAGYKLMIEIINQRPITIQDPLDQKCLDQVHGNIEFKDVTFSYSPRPDVMIFRNFNIFFPSPGKTAVVVGSGS
atpgp1	298	RHHLTGGLATAMTFVMIGGLALGQSAPMAAFKAKVAAAKIFRIIDHKPTIERNSESGVELDSTVGLVELKNVDFSPSPRPDVKILNNTFCLSVPAKTIALVGS
atpgp2	291	HKDIADGGKSTTMLNVVIAGLSLQAAPDISAFVRAKAAAVPIFKMIERNVTVTKTSKSGRKLGVGDGHIQFKDATFSYSPRPDVVIFDRNLAIIPAGKIVALVGS
consensus	331	is eytiG antvffsiligafsvGqaap idAFanargAay ifkiidn psidsfs Ghkpd ikGnlefkdvHFsYPSR evkilkgnlnkv sgqtvalVG SGC

W_A

FIGURE 1

hmdr3 434 GKSTTVQLIQRLYDPDEGTINIDGQDIRNFNVNRYLREIIGVVSQEPVLFSTTIAENICYGRNVMTDEIKKAVKEANAYFIMKLPKQKFDTLVGERGAQLSGGQKQRIAI
hmdr2 431 GKSTTVQLLQRLYDPTEGKISIDGQDIRNFNVNRYLREIIGVVSQEPVLFSTTIAENICYGRNVMTDEIKKAVKEANAYFIMKLPKQKFDTLVGERGAQLSGGQKQRIAI
hmdr1 432 GKSTTVQLMQRLYDPTEGMVSDGQDIRNFNVNRYLREIIGVVSQEPVLFSTTIAENICYGRNVMTDEIKKAVKEANAYFIMKLPKQKFDTLVGERGAQLSGGQKQRIAI
nmdr1 431 GKSTTVQLMQRLYDPTEGMVSDGQDIRNFNVNRYLREIIGVVSQEPVLFSTTIAENICYGRNVMTDEIKKAVKEANAYFIMKLPKQKFDTLVGERGAQLSGGQKQRIAI
atpac 405 GKSTTVSLIERFYDPNSGOILLDGVKIKTLQKFLRQIGLVNQEPALFATITILENILEYKFDATVFEAAASANAHSFITLLPKGYDTQVGERGLQSGGQKQRIAI
atpgp1 408 GKSTTVSLIERFYDPNSGOILLDGVKIKTLQKFLRQIGLVNQEPALFATITILENILEYKFDATVFEAAASANAHSFITLLPKGYDTQVGERGLQSGGQKQRIAI
atpgp2 401 GKSTTVSLIERFYEPISGAVLLDGNISLSELDIKWLRLQIGLVNQEPALFATITIRENILEYKFDATVFEAAASANAHSFITLLPKGYDTQVGERGLQSGGQKQRIAI
consensus 441 GKSTTVQLIQRLYDP EG V IDGQDIRNFNVNRYLREIIGVVSQEPVLFSTTIAENI YGR DVTMDIEKAVKEANAYFIMKLP fdtIVGeRGaQLSGGQKQRIAI

hmdr3 544 ARALVRNPKILLDEATSAIDTESAEVQAALDKAREGRTTIVIAHRLSTVRNADVIAGFEDGVIVEQSGHSELMKK. .EGVYFKLVNMQTSQIQSEE.FE
hmdr2 541 ARALVRNPKILLDEATSAIDTESAEVQAALDKAREGRTTIVIAHRLSTVRNADVIAGFEDGVIVEQSGHSELMKK. .EGYFRLVNMQTAGSILSEE.FE
hmdr1 542 ARALVRNPKILLDEATSAIDTESAEVQAALDKAREGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKE. .KGIYFKLVMTQTAGNEVELEN.AA
nmdr1 541 ARALVRNPKILLDEATSAIDTESAEVQAALDKAREGRTTIVIAHRLSTVRNADVIAGFDDGVIVEQSGHSELMKK. .KGIYFKLVMTQTAGNEIEPGN.NA
atpac 515 ARALVNDPKILLDEATSAIDTESAEVQAALDKAREGRTTIVIAHRLSTVRNADVIAGFDDGVIVEQSGHSELMKK. .SGAYASLIRFQEMVTRDFNSPSTRSTR
atpgp1 518 ARALVNDPKILLDEATSAIDTESAEVQAALDKAREGRTTIVIAHRLSTVRNADVIAGFDDGVIVEQSGHSELMKK. .SGAYASLIRFQEMVTRDFNSPSTRSTR
atpgp2 511 SRALVNDPKILLDEATSAIDTESAEVQAALDKAREGRTTIVIAHRLSTVRNADVIAGFDDGVIVEQSGHSELMKK. .SGAYASLIRFQEMVTRDFNSPSTRSTR
consensus 551 ARALVRNPKILLDEATSAIDTESAEVQAALDKAREGRTTIVIAHRLSTVRNADVIAGFDDGVIVEQSGHSELMKK Gvyfklv mqt g i n

hmdr3 643 .ELNDEKAATRMAPNGWKSRLFRHSTQKNLKNQMCOK. . .SLDVEDGLEANVPVFLKVLKLNKTEWPFYVGTVCALANGGLOPASFVIFSEIIAIFGPGDD. AVK
hmdr2 641 VELSDKAAGDVAPNGWKSRLFRHSTQKNLKNQMCOK. . .SLDVEDGLEANVPVFLKVLKLNKTEWPFYVGTVCALANGGLOPASFVIFSEIIAIFGPGDD. AVK
hmdr1 642 DESKSEIDALEMSNDSSRLIRKSTRRSVRGSOAD. . .RLDEETNEIDANVPVFLKVLKLNKTEWPFYVGTVCALANGGLOPASFVIFSEIIAIFGPGDD. AVK
nmdr1 641 YGQSQDFTASELTSEESKSLIR. RSIYRSVRKQDOE. . .RLSMKEAVDEDPVLSVFWIRILNLSWEPYLVGVLCVINGCIGQPVFAIVFSRIVGVFSRDDHETK
atpac 623 LSHSLFKSL.RGSRLNLSYSYTGADGRIEMISNAETDRKTRA. . .PENYFRLKLNSEWPYSYMGAVGSIISGFIPTFAIVMSNMIEVYTYTDYSME
atpgp1 628 ARNSVSSPIMTRNSSYGRSPYRRLSDFTSDFSLSDASSYPNRYNEKLAFOOANSFWRLAKMNSPEWKYALIGSVGICSAFFAYVLSAVLSVYVYNNPDHEYMI
atpgp2 620 IKYS.RELSTRSSFCSE. ESVTRPDGADPSKKVKTVG. . .RLYSMIRPDMMYGVGCTICAFIAGSQMPLFALGVSQAL.VSYYSGWDETQ
consensus 661 s e a m ks l R s s qd r d le vp vsfwrvlkn tewpy vvgtvcaing lqp fails iiafv dd vk

hmdr3 748 QQKCNIFSLIFLGLISFTFFLOGFTFGKAGEILTRRLRSMAFKAMLRQDMSWFDHDKNSTGALSTRLATDAAQVQATGTRLALIAQNIANLGTGIIISFIYGWQLT
hmdr2 745 QQKCNMFSLVFLGLVLSFTFFLOGFTFGKAGEILTRRLRSMAFKAMLRQDMSWFDHDKNSTGALSTRLATDAAQVQATGTRLALIAQNIANLGTGIIISFIYGWQLT
hmdr1 749 QNSNLSLFLFALGIIISFTFFLOGFTFGKAGEILTRRLRSMAFKAMLRQDMSWFDHDKNSTGALSTRLATDAAQVQATGTRLALIAQNIANLGTGIIISFIYGWQLT
nmdr1 747 QNCNLSLFFVYGLISFTFFLOGFTFGKAGEILTRRVYVFKSMLRQDISWFDHDKNSTGALSTRLATDAAQVQATGTRLALIAQNIANLGTGIIISFIYGWQLT
atpac 725 RK.TKEYVFIYIGAGLYAVGAYLIHYFFSINGENLTTRVRMMLSAIRNEVGEDEHNSSLIAARLADANNRVSALIAERISVLQNMTSLLTSFTVAFIVEWRVS
atpgp1 738 QK.IDKCYLLIGLSSAALVENTLQHSFWDIVGENLTTRVRMMLSAIRNEVGEDEHNSSLIAARLADANNRVSALIAERISVLQNMTSLLTSFTVAFIVEWRVS
atpgp2 707 KE.IKKTAILFCCASVITLIVTIBHCFTGMEGRLTRVRMMLSAIRNEVGEDEHNSSLIAARLADANNRVSALIAERISVLQNMTSLLTSFTVAFIVEWRVS
consensus 771 rq nifsliflglgiisfitflggftfgkageiltrvr mvfkamLRqdmSWFD kntg 1stRlatDaaqvkgag rla vi QNianlgtgiisfiywgwlt

FIGURE 2

858 LLLAVVPPIAVSGIVEMKLIAGNAKRDKKELEAAGKIATEAIENTRTVVSILTQERKFESMYVEKLYGPRNSV..QKAHIXGITEFISQAEMFYSYAGCFRFGAYLIVN
855 LLLSVVPPIAVAGIVEMKMLAGNAKRDKKEAAGKIATEAIENTRTVVSILTQERKFESMYVEKLGHPYNSV..RKAHIXGITEFISQAEMFYSYAGCFRFGAYLIVN
859 LLLAIVPIIAIAGVEMKMLSGQALKDKKELEGAGKIATEAIENTRTVVSILTQEQFEHMYAQSLQVPYNSL..RKAHIGITESTTQAMMYFSYAGCFRFGAYLVAH
857 LLLVVIIPILVILGGIEMKLLSGQALKDKLEISGKIATEAIENTRTVVSILTREQFEHMYAQSLQVPYNSL..RKAHIGITESTTQAMMYFSYAGCFRFGAYLVAH
834 LLLILGTFLLVLNFAQQLSLKGFAGDKTAKAHAKTSMIAGEGVSNI RTVAAFNAQSLKLSFCHHELVPKRSLSIYRSQTSGLFGLSQALALYSEALIMYG AHLVSK
847 LILVAVFPVVAATVLOQMFMTGSGDLEAAHAKGTQAGEAIA NVRTVAAFNSEAKIVRLYXANLEPPIKR..CFWKGQIAGSGYGVQAQFCLYASVALGYASWLVKH
816 LVVLIATYPIVISGHI SEKLFMQGYGGDLNKAYLKANMLAGESVSNIRTVAAFCAEKILELYSRELLEPSKSS..FRGQIAGLFYGVSQFFIFSSYGLALWYGSTLMDK
consensus 881 Llllavvpilvvagivemkil Gna rdkk le agkiatEaienrtvvsilt e kfesmy L Pynsv rkahiyGittfsisQa myfSya gcfrrfgaylv h

W_A

966 GHMFRDVIILVFSaIVFGAVALGHASFPADYAKAKLSAAHLFMLFERQPLIDSYSSEGL.KPDKFEGNITFNEVVFNYPTRANVPVLQGLSLEVKKGQTILALVGS GCG
963 GHMRFKDVILVFSaIVLGAVALGHASFPADYAKAKLSAAYLFSLFERQPLIDSYSSEGL.WPDKFEGSVTFNEVVFNYPTRANVPVLQGLSLEVKKGQTILALVGS GCG
967 KLMSFEDVILVFSaVFGAMAVGQVSSFPADYAKAKLSAAHIIIMIEKTPLIDSYSSEGL.MENTLEGNVTFGEVFNYPTRPDIPVLQGLSLEVKKGQTILALVGS GCG
965 QLMTFENVMLVFSaVFGAMAAGNTSSFPADYAKAKVSASHIIRIEKTEPLIDSYSSEGL.KPTLLEGNVKFNQFNYPTRPNIPVLQGLSLEVKKGQTILALVGS GCG
944 GVSTFSKVIKVFVVLVITANSVAETVSLAPEIIRGGEAVGSVSVLDRQTRIDPDADADPV.ETIRGDIIEFRHVDFAYPSPDPVDFNLRIRAGHSOALVGA SGG
955 GISDFSKTIRVFMVIMVSANGAETILTAPDFIKGGQAMRSVFELDRKTEIEPDDPTTPVDRLRGEVELKHIDFSYSPRPDIQIFRDLISLRARAGKTILALVGS GCG
924 GLAGFKSVMTFMVLIVTALAMGETILALAPDLLKGNQWVASVFEILDRKTOIV...GETSEELNNVEGTIELKGVHFSYSPRPDVIFRDFDLIVRAGKSMALVGS GSG
consensus 991 glm F vilvFsaivlgavalg tssfApdyakaklsaa lf hier p ldsys egl pd leG v f v FnyPcRpdvpvliqglsl evkkkgqtlALVGS GCG

1075 KSTVVQLLERFYDPLAGTVLLDGOBAKKNVQWLRQAOLGIVSQEPILFDCSIAENIAYGDNRSRVVSQDEIVSAAKAAIHPIETLPHKYETRVGDKGTOLSGGQKORIA
1072 KSTVVQLLERFYDPMAGSVLLDGOBAKKNVQWLRQAOLGIVSQEPILFDCSIAENIAYGDNRSRVVPHDEIVRAAKEANIHPIETLPKYNTRVGDKGTOLSGGQKORIA
1076 KSTVVQLLERFYDPLAGKVLLDGGKTKRLNVQWLRQAOLGIVSQEPILFDCSIAENIAYGDNRSRVVSQDEIVRAAKEANIHAFIESLPNKYSTRVGDKGTOLSGGQKORIA
1074 KSTVVQLLERFYDPMAGSVLLDGGKTKQLNVQWLRQAOLGIVSQEPILFDCSIAENIAYGDNRSRVVSHDEIVRAAKEANIHQFIDSLPKYNTRVGDKGTOLSGGQKORIA
1053 KSSVIAMERFYDLAGKVMIDGKDRLRLNLKSLRLKTLGVQEPALFAATI FDNIA YGKDG..ATESEVIDAARAANAHAHGFISGLPEGYKTPVGERGVQLSGGQKORIA
1065 KSSVISLIQRFYEPSSGRVMDGKDRLKYNLKAIRKHIAIVQEPCLFGTTIYENIAYGHEC..ATEAEIIQAATLASAHKFI SALPEGYKTYPVGERGVQLSGGQKORIA
1031 KSSVISLILRFYDPTAGKVMIEGKDIKKLDLKAIRKHIAIVQEPALFAATTIYENIAYNEG..ASQSEVVESAMLANAHAFITSLPEGYSTKYPVGERGVQLSGGQKORIA
consensus 1101 KstVvqlleRfYdplagkvllldgkeikklnvqwlRahligiVsQEPiLfdcsIaeNiaYgdnsr vs dEiv aak AniH FietLPdky TrvcdkgtQlsggQkORIA

FIGURE 3

W_B

hmdr3	1185	IARALIROPQIILLDEATSEKVVQAEALDKAREGTCIVIAHRLSTIQNADLIWVFQNGRVKKEHGHQQLAQK..GIYFSMVSVOAGTONL~~~~~
mmdr2	1182	IARALIROPRLVLLDEATSEKVVQAEALDKAREGTCIVIAHRLSTIQNADLIWVFQNGRVKKEHGHQQLAQK..GIYFSMNIQAGTONL~~~~~
hmdr1	1186	IARALVROPHIILLDEATSEKVVQAEALDKAREGTCIVIAHRLSTIQNADLIWVFQNGRVKKEHGHQQLAQK..GIYFSMVSVOAGTKRQ~~~~~
mmdr1	1184	IARALVROPHIILLDEATSEKVVQAEALDKAREGTCIVIAHRLSTIQNADLIWVFQNGRVKKEHGHQQLAQK..GIYFSM..VQAGAKRS~~~~~
atpac	1161	IARAVLKNPTVLLDEATSEKVVQAEALDKAREGTCIVIAHRLSTIRGVDCIGVIOGGRIVEOGSHSELV.SRPEGAYSRLQLQTHRI*~~~~~
atpgp1	1173	IARALVRKAEIMLLDEATSEKVVQAEALDKAREGTCIVIAHRLSTIRNAHVIAVIDDGKVAEQGSHSHLLKNHPDGIYARMIQLQRFTHTQVIGMTSGSSSRVK
atpgp2	1139	IARAILKNPAIILLDEATSEKVVQAEALDKAREGTCIVIAHRLSTIKNADTISVLHGGKIVEOGSHRKLVLNK.SGPYFKLISLQQQQQP~~~~~
consensus	1211	IARALirqp iLLDEATSEKvvQeALdkaregtCivIAHRLSTiqnadliWv i nckvkEhGtHqqLlaqk GiYfsmv vQagt
hmdr3	1280	~~~~~
mmdr2	1277	~~~~~
hmdr1	1281	~~~~~
mmdr1	1277	~~~~~
atpac	1255	~~~~~
atpgp1	1283	EDDA
atpgp2	1234	~~~~~
consensus	1321	

FIGURE 4.

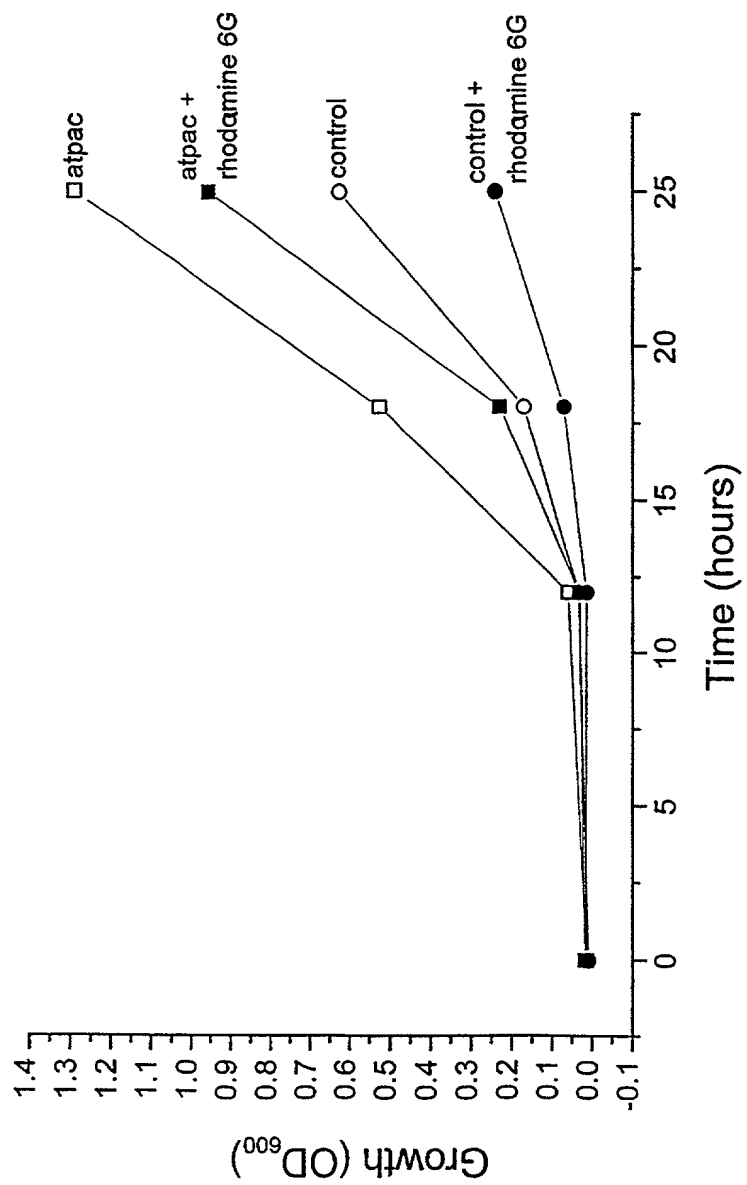


FIGURE 5

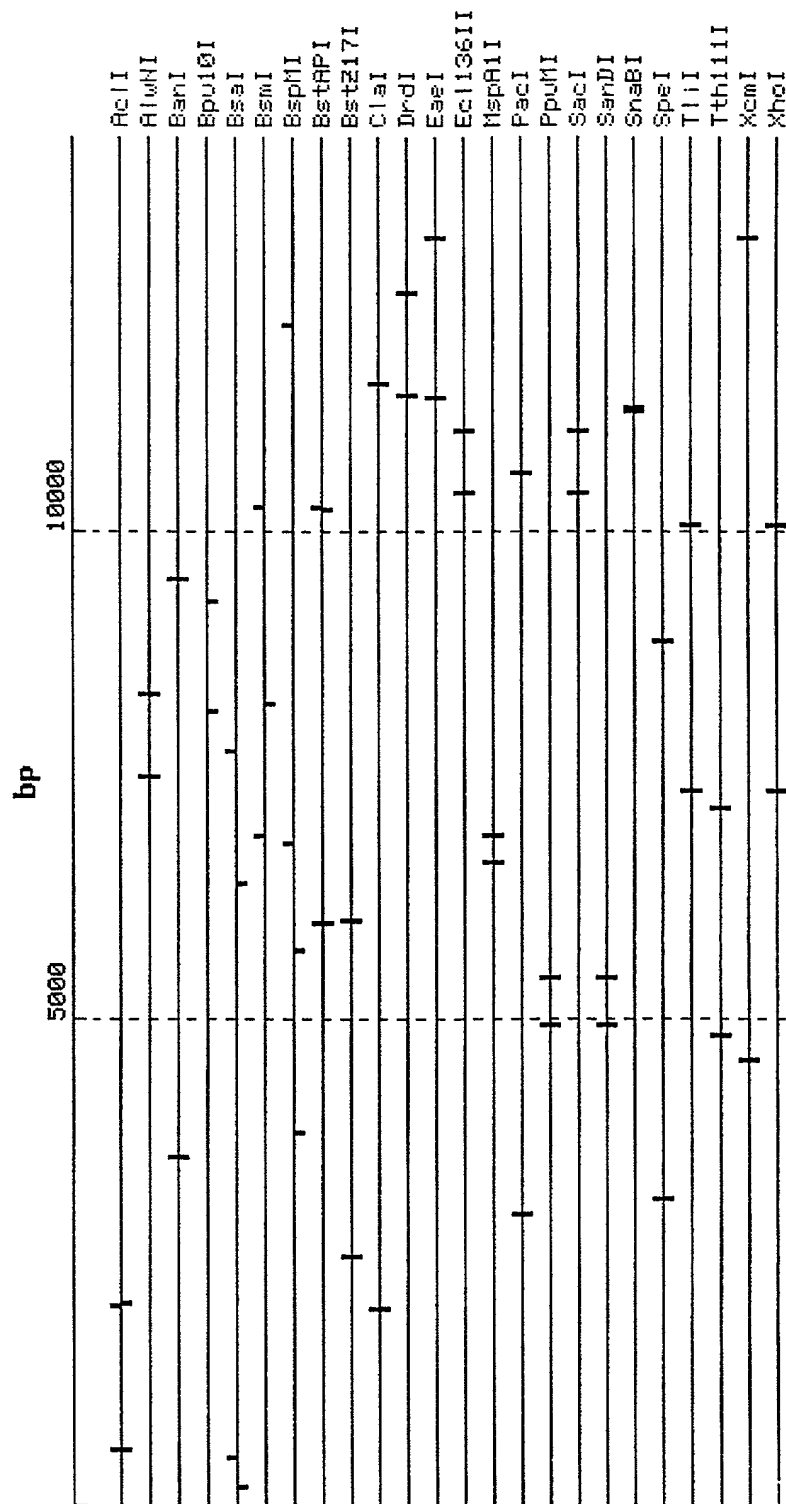


FIGURE 6

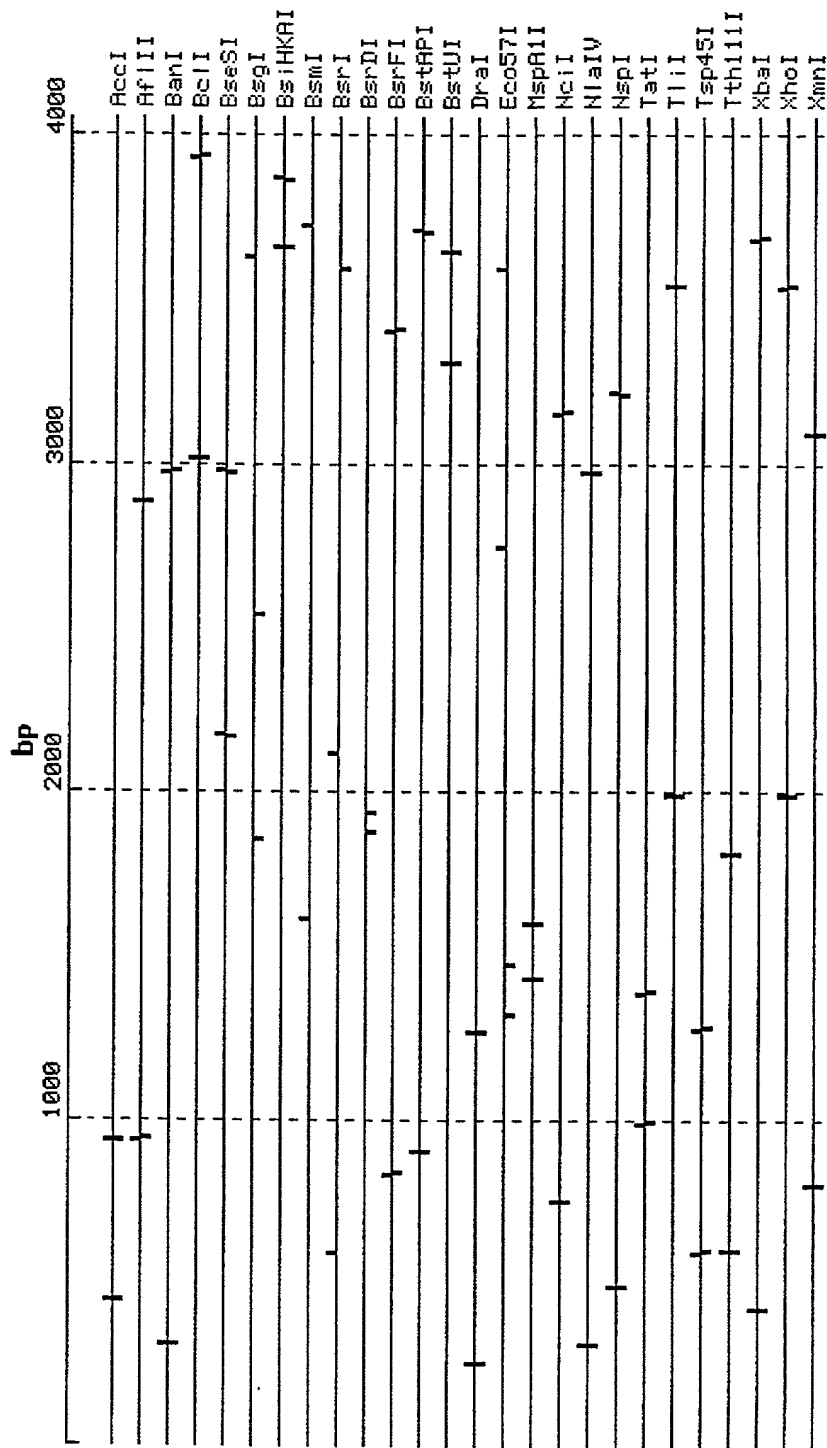


FIGURE 7

6

Sequence Listing



SEQUENCE LISTING

<110> Spalding, Edgar P.
Noh, Bosl

<120> MDR-Like ABC Transporter Gene From
Plants

<130> 13238-00061

<140> herewith
<141> not yet assigned

<150> PCT/US99/22363
<151> 1999-09-24

<150> US 60/101,814
<151> 1998-09-25

<160> 14

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 4051
<212> DNA
<213> Arabidopsis thaliana

<220>

<221> misc_feature
<222> (94) .. (0)

<223> Translation start codon

<221> misc_feature

<222> (3932) .. (0)
<223> Stop codon

<400> 1

cttgaactc acaaaacaat tgtcagattt tcaagaaaaa cttataaaa caaaaaacat
ttcattctt ctcctctct ctcctcactgc tcaatgatct cgtttctca ctaaaccaac
tcgttcttc ttaattctt taactcggat ctacaaaaa ccatgtcgg aactaacaca
accgatgcc agactgtcc agcagaagca gagagaagca agagaacagag tttaaccatc
tttaacctc ttctctcgc tgataaatc gatlatctc taatgttcgt tggttctct
ggtgccattg gtaaaaaa aatgattta catcaatgg ttcattagat ctctagatat
tctctatat tctctactt gggtttggtc gtttgcttc ctctcacc agagatagca
tgttggtgt attctggaga aagaacaagt gacgactta ggaagaaata tcttgaaagca
gtattaaaac aagaagttgg gtctcttgat actgactta gaaactggtag catgtctt
agtttctc ctgatactc tcttgttcaa gatgccata gtgaaaaagt tggaaaactt
atacatcac tctcaacat ttggcggga ttagtagtg gatltgtac agcatggaaa
tagctttgt taagtgtgc tgtgattccc ggaatcgct tgcgcggagg ttatacgt
tatacactc ccggaattac tcaagaagc cgtgaatct atgctaagc cgttgtatc
gccagcagg caatgcaca agtcgaact gttatctc atgttgagga gagttaaggca
cctaagcgt attcggatgc gattcagtat acccttaagc tcggtataa agcggggtg
gctaaagggt tgggttagg atgtacttat ggaatagct gtatgtcatg ggcgttggtg
tttgglatg ctgagtttt tattcggaat ggaacaaccg atggaggaag ggcgtttact

60
120
180
240
300
360
420
480
540
600
660
720
780
840
900
960
1020
1080

SEQUENCE LISTING

<110> Spalding, Edgar P.
Noh, Bosl

<120> MDR-Like ABC Transporter Gene From
Plants

<130> 13238-00061

<140> herewith

<141> not yet assigned

<150> PCT/US99/22363

<151> 1999-09-24

<150> US 60/101,814

<151> 1998-09-25

<160> 14

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 4051

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> misc_feature

<222> (94)...(0)

<223> Translation start codon

<221> misc_feature

<222> (3932)...(0)

<223> Stop codon

<400> 1

cttgaacttc	acaaaacaat	tgctcagattt	tcaagaaaaa	ctttataaaa	caaaaaaacat	60
ttcattcttt	ctctctctct	ctctcactgc	tcaatgatct	cgttttctca	ctaaaccaac	120
tcgtttcttc	ttactttctt	taactcggat	ctacaaaaaa	ccatgtcggg	aactaacaca	180
accgatgcc	agactgttcc	agcagaagca	gagaagaaga	aagaacagag	tttaccattc	240
tttaaaactc	tttcttttgc	tgataaaatt	gattatctct	taatgttcgt	tggttctctt	300
ggtgccattg	ttcatggctc	ttccatgcct	gtcttctttt	tactctttgg	tcaaattggt	360
aatggatttg	gtaaaaacca	aatggattta	catcaaattg	ttcatgaagt	ctctagatat	420
tctctatatt	tcgtctactt	gggttttggt	gtttgcttct	cttcttacgc	agagatagca	480
tgttggatgt	attctggaga	aagacaagta	gcagcattaa	ggaagaaata	tcttgaagca	540
gtattaaaac	aagacgttgg	gttctttgat	actgatgcta	gaactggtga	cattgtcttt	600
agtgtttcta	ctgatactct	tcttgttcaa	gatgccatta	gtgaaaagg	tggaaacttt	660
atacattacc	tctcaacatt	tttggcggga	ttagtagttg	gatttgtatc	agcatggaaa	720
ttagctttgt	taagtgttgc	tgtgattccc	ggaatcgctt	tcgccggagg	tttatacgct	780
tatacactca	ccggaattac	ttcaaagagc	cgtgaatctt	atgctaacgc	cggtgttatc	840
gccgagcagg	caattgctca	agttcgaact	gtttattctt	atgttggaga	gagtaaggca	900
cttaatgcgt	attcggatgc	gattcagtat	acgcttaagc	tcggttataa	agcggggatg	960
gctaaaagggt	tgggttttag	atgtacttat	ggaatagctt	gtatgtcatg	ggctttggtg	1020
ttttggatatg	ctggagtttt	tattcggaat	ggacaaaccg	atggaggaaa	ggcgtttact	1080

gctatattct ctgctattgt tgggtggaatg agtttggggc aatctttctc gaatcttggg 1140
gcgttttagta aaggtaaaagc ggctgggttat aagttgatgg agataattaa ccagagaccg 1200
acgataattc aagacccgtt ggatggaaaaa tgtttggatc aagttcatgg gaacattgag 1260
tttaaagatg tgacttttag ctatccttca cggcctgatg ttatgatctt caggaacttt 1320
aatattttct tcccttctgg gaaaactgtg gcggttggtg gtgggagtgg ctctggaaag 1380
agtactgttg tttccctcat tgagagattc tatgatccaa acagcgggca aattctgttg 1440
gatgggtgtg agataaagac gcttcagttg aagtttttgc gtgaacaaat cgggcttgtg 1500
aatcaagaac ctgcgctctt tgccactact atactagaga acatactcta tggaaagcct 1560
gatgcaacaa tggttgaagt tgaagctgct gcttcgcgtg cgaatgcgca tagtttcatt 1620
acattacttc ctaaaggcta cgacacacag gttggagaac gtggtgttca actctcaggt 1680
ggacagaagc agagaattgc aattgctagg gogatgttga aagacccaaa gattctgtta 1740
ctagatgaag ctacaagcgc tcttgatgct agctctgaga gcattgttca ggaagcttta 1800
gacagagtca tgggtggggag gaccactggt gttgttgctc atcgtctctg caccatcaga 1860
aatgttgatt ccattgccgt gatacagcaa ggccaagttg ttgaaaccgg aacacatgaa 1920
gaactcattg ccaaatccgg tgcctacgca tccctcatca ggtttcagga aatggttggg 1980
actcgagatt tctcaaaccg gtcaactcgt cgcactcgtt caaccctgtt gagccattca 2040
ctgtcaacga aatcactcag tttaagatca ggaagtttga ggaatctgag ctattcttac 2100
agcactggag ctgatggctg gatagagatg atttcaaagc cagagactga ccgaaagact 2160
cgtgcccctg aaaattactt ctacaggctt ctcaggctta attcaccgga atggccttac 2220
tcaatcatgg gagcagtagg ctcaattctt tctggtttca ttggtcctac atttgctatt 2280
gtgatgagca acatgatcga agtcttctac tacacagact atgattcaat ggaaaggaaa 2340
acaaaagagt atgtcttcat ctacattggt gctggtctct atgctgtggg tgcttatttg 2400
atccaacatt acttcttttag catcatggga gaaaacctca caacaagagt aagaagaatg 2460
atgctctcag ctatcttgag aaacgaagtt ggttggttcg atgaggatga acacaactca 2520
agcctgatcg ctgcacgttt agctactgat gcagcagatg ttaaatccgc tatagccgag 2580
agaatctcag taattctaca aaacatgact tcacttctca catccttcat agtcgccttc 2640
atagtagaat ggagagtctc acttctcatc ttaggcacat tcccacttct agtcctcgct 2700
aactttgctc agcaactatc tctgaagggt tttgctggag acacagctaa ggctcatgca 2760
aagacttcaa tgattgctgg tgaaggagtc agtaacatta gaaccgtagc agctttcaat 2820
gcacagagca agattctctc tttgttctgt catgagcttc gtgtacctca gaaaagaagc 2880
ttaagcttat accgaagtca aacctcgggt ttcctatttg gcctctcgca gcttgctctc 2940
tatggttctg aggttttaat tctctggtat ggtgcccacc ttgtgagtaa aggcgtgtca 3000
accttttcca aagtgatcaa agtgtttgtg gttttggtca ttactgcaa ctctgttgct 3060
gaaactgtca gtcttgctcc tgaaattatt cggggaggtg aagctgttgg ttcggttttc 3120
tcggtcttgg acaggcagac caggattgac cgggatgatg ctgatgctga tcccggtggag 3180
acgatccgtg gagacattga gtttaggcat gttgatttcg cttacccttc aagacccgac 3240
gtcatgggtt tcagggaact taacctcaga attcgagctg gacatagcca agctcttgtg 3300
ggcgcgagtg ggtcagggaa gagttctgta attgcgatga tcgagcgggt ttacgacctt 3360
cttgctggaa aagtcatgat tgatggcaaa gacatccgcc ggctaaacct gaaatctcta 3420
aggctcaaaa tcggtcttgt tcaacaagaa ccagctcttt tcgcagcaac gatcttcgac 3480
aacatcgctt atggttaaaga tgggtgcaact gaatccgagg taattgatgc agctcgagcc 3540
gcaaatgctc acggtttcat cagtggttta cctgaagggt acaaaaactcc agtaggcgaa 3600
agaggagtgc agttatcagg tggacagaaa cagaggatcg cgatagcaag agctgtgctc 3660
aagaacccta cagtgttgct tctagacgaa gcaactagcg cactagatgc agaatacagaa 3720
tgcgtgctgc aagaggcggt agagaggctc atgagaggtc ggaccaccgt ggtagttgct 3780
caccgcttgt ccaccataag aggtgttgat tgcattggtg tgattcaaga cgggcggatt 3840
gtggagcaag gcagccattc agagctcgtt agccgaccag agggagctta ttcaaggctg 3900
ttacagcttc aaacacatag gatttgaagc ttgatcatgg attaaaaaca aaaaatcggt 3960
ttgtgtaatt ttttttatat taaaacttta atttggaaga tttctatgga ctataacgat 4020
aatatgaata ggtgtagata atgaagcttt t 4051

<210> 2
<211> 1254
<212> PRT
<213> Arabidopsis thaliana

<400> 2

Met	Ser	Glu	Thr	Asn	Thr	Thr	Asp	Ala	Lys	Thr	Val	Pro	Ala	Glu	Ala
1				5					10					15	
Glu	Lys	Lys	Lys	Glu	Gln	Ser	Leu	Pro	Phe	Phe	Lys	Leu	Phe	Ser	Phe
			20					25					30		
Ala	Asp	Lys	Phe	Asp	Tyr	Leu	Leu	Met	Phe	Val	Gly	Ser	Leu	Gly	Ala
		35				40					45				
Ile	Val	His	Gly	Ser	Ser	Met	Pro	Val	Phe	Phe	Leu	Leu	Phe	Gly	Gln
50					55					60					
Met	Val	Asn	Gly	Phe	Gly	Lys	Asn	Gln	Met	Asp	Leu	His	Gln	Met	Val
65					70					75				80	
His	Glu	Val	Ser	Arg	Tyr	Ser	Leu	Tyr	Phe	Val	Tyr	Leu	Gly	Leu	Val
				85					90					95	
Val	Cys	Phe	Ser	Ser	Tyr	Ala	Glu	Ile	Ala	Cys	Trp	Met	Tyr	Ser	Gly
			100					105					110		
Glu	Arg	Gln	Val	Ala	Ala	Leu	Arg	Lys	Lys	Tyr	Leu	Glu	Ala	Val	Leu
		115					120						125		
Lys	Gln	Asp	Val	Gly	Phe	Phe	Asp	Thr	Asp	Ala	Arg	Thr	Gly	Asp	Ile
130						135					140				
Val	Phe	Ser	Val	Ser	Thr	Asp	Thr	Leu	Leu	Val	Gln	Asp	Ala	Ile	Ser
145					150					155					160
Glu	Lys	Val	Gly	Asn	Phe	Ile	His	Tyr	Leu	Ser	Thr	Phe	Leu	Ala	Gly
				165					170					175	
Leu	Val	Val	Gly	Phe	Val	Ser	Ala	Trp	Lys	Leu	Ala	Leu	Leu	Ser	Val
			180					185					190		
Ala	Val	Ile	Pro	Gly	Ile	Ala	Phe	Ala	Gly	Gly	Leu	Tyr	Ala	Tyr	Thr
		195					200						205		
Leu	Thr	Gly	Ile	Thr	Ser	Lys	Ser	Arg	Glu	Ser	Tyr	Ala	Asn	Ala	Gly
210						215						220			
Val	Ile	Ala	Glu	Gln	Ala	Ile	Ala	Gln	Val	Arg	Thr	Val	Tyr	Ser	Tyr
225					230					235					240
Val	Gly	Glu	Ser	Lys	Ala	Leu	Asn	Ala	Tyr	Ser	Asp	Ala	Ile	Gln	Tyr
				245					250					255	
Thr	Leu	Lys	Leu	Gly	Tyr	Lys	Ala	Gly	Met	Ala	Lys	Gly	Leu	Gly	Leu
			260					265					270		
Gly	Cys	Thr	Tyr	Gly	Ile	Ala	Cys	Met	Ser	Trp	Ala	Leu	Val	Phe	Trp
		275					280						285		
Tyr	Ala	Gly	Val	Phe	Ile	Arg	Asn	Gly	Gln	Thr	Asp	Gly	Gly	Lys	Ala
290						295					300				
Phe	Thr	Ala	Ile	Phe	Ser	Ala	Ile	Val	Gly	Gly	Met	Ser	Leu	Gly	Gln
305					310					315					320
Ser	Phe	Ser	Asn	Leu	Gly	Ala	Phe	Ser	Lys	Gly	Lys	Ala	Ala	Gly	Tyr
				325					330					335	
Lys	Leu	Met	Glu	Ile	Ile	Asn	Gln	Arg	Pro	Thr	Ile	Ile	Gln	Asp	Pro
			340					345					350		
Leu	Asp	Gly	Lys	Cys	Leu	Asp	Gln	Val	His	Gly	Asn	Ile	Glu	Phe	Lys
		355					360					365			
Asp	Val	Thr	Phe	Ser	Tyr	Pro	Ser	Arg	Pro	Asp	Val	Met	Ile	Phe	Arg
370						375					380				
Asn	Phe	Asn	Ile	Phe	Phe	Pro	Ser	Gly	Lys	Thr	Val	Ala	Val	Val	Gly
385					390					395					400
Gly	Ser	Gly	Ser	Gly	Lys	Ser	Thr	Val	Val	Ser	Leu	Ile	Glu	Arg	Phe
				405					410					415	
Tyr	Asp	Pro	Asn	Ser	Gly	Gln	Ile	Leu	Leu	Asp	Gly	Val	Glu	Ile	Lys
			420					425					430		
Thr	Leu	Gln	Leu	Lys	Phe	Leu	Arg	Glu	Gln	Ile	Gly	Leu	Val	Asn	Gln
		435					440					445			
Glu	Pro	Ala	Leu	Phe	Ala	Thr	Thr	Ile	Leu	Glu	Asn	Ile	Leu	Tyr	Gly

450		455		460
Lys Pro Asp Ala Thr Met Val Glu Val Glu Ala Ala Ala Ser Ala Ala				
465		470		475
Asn Ala His Ser Phe Ile Thr Leu Leu Pro Lys Gly Tyr Asp Thr Gln				480
	485		490	495
Val Gly Glu Arg Gly Val Gln Leu Ser Gly Gly Gln Lys Gln Arg Ile				
	500		505	510
Ala Ile Ala Arg Ala Met Leu Lys Asp Pro Lys Ile Leu Leu Leu Asp				
	515		520	525
Glu Ala Thr Ser Ala Leu Asp Ala Ser Ser Glu Ser Ile Val Gln Glu				
	530		535	540
Ala Leu Asp Arg Val Met Val Gly Arg Thr Thr Val Val Val Ala His				
545		550		555
Arg Leu Cys Thr Ile Arg Asn Val Asp Ser Ile Ala Val Ile Gln Gln				560
	565		570	575
Gly Gln Val Val Glu Thr Gly Thr His Glu Glu Leu Ile Ala Lys Ser				
	580		585	590
Gly Ala Tyr Ala Ser Leu Ile Arg Phe Gln Glu Met Val Gly Thr Arg				
	595		600	605
Asp Phe Ser Asn Pro Ser Thr Arg Arg Thr Arg Ser Thr Arg Leu Ser				
	610		615	620
His Ser Leu Ser Thr Lys Ser Leu Ser Leu Arg Ser Gly Ser Leu Arg				
625		630		635
Asn Leu Ser Tyr Ser Tyr Ser Thr Gly Ala Asp Gly Arg Ile Glu Met				
	645		650	655
Ile Ser Asn Ala Glu Thr Asp Arg Lys Thr Arg Ala Pro Glu Asn Tyr				
	660		665	670
Phe Tyr Arg Leu Leu Lys Leu Asn Ser Pro Glu Trp Pro Tyr Ser Ile				
	675		680	685
Met Gly Ala Val Gly Ser Ile Leu Ser Gly Phe Ile Gly Pro Thr Phe				
	690		695	700
Ala Ile Val Met Ser Asn Met Ile Glu Val Phe Tyr Tyr Thr Asp Tyr				
705		710		715
Asp Ser Met Glu Arg Lys Thr Lys Glu Tyr Val Phe Ile Tyr Ile Gly				
	725		730	735
Ala Gly Leu Tyr Ala Val Gly Ala Tyr Leu Ile Gln His Tyr Phe Phe				
	740		745	750
Ser Ile Met Gly Glu Asn Leu Thr Thr Arg Val Arg Arg Met Met Leu				
	755		760	765
Ser Ala Ile Leu Arg Asn Glu Val Gly Trp Phe Asp Glu Asp Glu His				
	770		775	780
Asn Ser Ser Leu Ile Ala Ala Arg Leu Ala Thr Asp Ala Ala Asp Val				
785		790		795
Lys Ser Ala Ile Ala Glu Arg Ile Ser Val Ile Leu Gln Asn Met Thr				
	805		810	815
Ser Leu Leu Thr Ser Phe Ile Val Ala Phe Ile Val Glu Trp Arg Val				
	820		825	830
Ser Leu Leu Ile Leu Gly Thr Phe Pro Leu Leu Val Leu Ala Asn Phe				
	835		840	845
Ala Gln Gln Leu Ser Leu Lys Gly Phe Ala Gly Asp Thr Ala Lys Ala				
	850		855	860
His Ala Lys Thr Ser Met Ile Ala Gly Glu Gly Val Ser Asn Ile Arg				
865		870		875
Thr Val Ala Ala Phe Asn Ala Gln Ser Lys Ile Leu Ser Leu Phe Cys				
	885		890	895
His Glu Leu Arg Val Pro Gln Lys Arg Ser Leu Ser Leu Tyr Arg Ser				
	900		905	910

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Gln Thr Ser Gly Phe Leu Phe Gly Leu Ser Gln Leu Ala Leu Tyr Gly
 915 920 925
 Ser Glu Ala Leu Ile Leu Trp Tyr Gly Ala His Leu Val Ser Lys Gly
 930 935 940
 Val Ser Thr Phe Ser Lys Val Ile Lys Val Phe Val Val Leu Val Ile
 945 950 955 960
 Thr Ala Asn Ser Val Ala Glu Thr Val Ser Leu Ala Pro Glu Ile Ile
 965 970 975
 Arg Gly Gly Glu Ala Val Gly Ser Val Phe Ser Val Leu Asp Arg Gln
 980 985 990
 Thr Arg Ile Asp Pro Asp Asp Ala Asp Ala Asp Pro Val Glu Thr Ile
 995 1000 1005
 Arg Gly Asp Ile Glu Phe Arg His Val Asp Phe Ala Tyr Pro Ser Arg
 1010 1015 1020
 Pro Asp Val Met Val Phe Arg Asp Phe Asn Leu Arg Ile Arg Ala Gly
 1025 1030 1035 1040
 His Ser Gln Ala Leu Val Gly Ala Ser Gly Ser Gly Lys Ser Ser Val
 1045 1050 1055
 Ile Ala Met Ile Glu Arg Phe Tyr Asp Leu Leu Ala Gly Lys Val Met
 1060 1065 1070
 Ile Asp Gly Lys Asp Ile Arg Arg Leu Asn Leu Lys Ser Leu Arg Leu
 1075 1080 1085
 Lys Ile Gly Leu Val Gln Gln Glu Pro Ala Leu Phe Ala Ala Thr Ile
 1090 1095 1100
 Phe Asp Asn Ile Ala Tyr Gly Lys Asp Gly Ala Thr Glu Ser Glu Val
 1105 1110 1115 1120
 Ile Asp Ala Ala Arg Ala Ala Asn Ala His Gly Phe Ile Ser Gly Leu
 1125 1130 1135
 Pro Glu Gly Tyr Lys Thr Pro Val Gly Glu Arg Gly Val Gln Leu Ser
 1140 1145 1150
 Gly Gly Gln Lys Gln Arg Ile Ala Ile Ala Arg Ala Val Leu Lys Asn
 1155 1160 1165
 Pro Thr Val Leu Leu Leu Asp Glu Ala Thr Ser Ala Leu Asp Ala Glu
 1170 1175 1180
 Ser Glu Cys Val Leu Gln Glu Ala Leu Glu Arg Leu Met Arg Gly Arg
 1185 1190 1195 1200
 Thr Thr Val Val Val Ala His Arg Leu Ser Thr Ile Arg Gly Val Asp
 1205 1210 1215
 Cys Ile Gly Val Ile Gln Asp Gly Arg Ile Val Glu Gln Gly Ser His
 1220 1225 1230
 Ser Glu Leu Val Ser Arg Pro Glu Gly Ala Tyr Ser Arg Leu Leu Gln
 1235 1240 1245
 Leu Gln Thr His Arg Ile
 1250

<210> 3
 <211> 1280
 <212> PRT
 <213> Homo sapiens

<300>
 <308> Genbank P08183
 <309> 1997-11-01

<400> 3

Met Asp Leu Glu Gly Asp Arg Asn Gly Gly Ala Lys Lys Lys Asn Phe
 1 5 10 15

Phe	Lys	Leu	Asn	Asn	Lys	Ser	Glu	Lys	Asp	Lys	Lys	Glu	Lys	Lys	Pro			
			20					25					30					
Thr	Val	Ser	Val	Phe	Ser	Met	Phe	Arg	Tyr	Ser	Asn	Trp	Leu	Asp	Lys			
		35					40					45						
Leu	Tyr	Met	Val	Val	Gly	Thr	Leu	Ala	Ala	Ile	Ile	His	Gly	Ala	Gly			
	50					55					60							
Leu	Pro	Leu	Met	Met	Leu	Val	Phe	Gly	Glu	Met	Thr	Asp	Ile	Phe	Ala			
65				70					75						80			
Asn	Ala	Gly	Asn	Leu	Glu	Asp	Leu	Met	Ser	Asn	Ile	Thr	Asn	Arg	Ser			
			85					90						95				
Asp	Ile	Asn	Asp	Thr	Gly	Phe	Phe	Met	Asn	Leu	Glu	Glu	Asp	Met	Thr			
		100						105					110					
Arg	Tyr	Ala	Tyr	Tyr	Tyr	Ser	Gly	Ile	Gly	Ala	Gly	Val	Leu	Val	Ala			
		115					120					125						
Ala	Tyr	Ile	Gln	Val	Ser	Phe	Trp	Cys	Leu	Ala	Ala	Gly	Arg	Gln	Ile			
	130					135					140							
His	Lys	Ile	Arg	Lys	Gln	Phe	Phe	His	Ala	Ile	Met	Arg	Gln	Glu	Ile			
145				150						155					160			
Gly	Trp	Phe	Asp	Val	His	Asp	Val	Gly	Glu	Leu	Asn	Thr	Arg	Leu	Thr			
			165					170						175				
Asp	Asp	Val	Ser	Lys	Ile	Asn	Glu	Val	Ile	Gly	Asp	Lys	Ile	Gly	Met			
		180						185					190					
Phe	Phe	Gln	Ser	Met	Ala	Thr	Phe	Phe	Thr	Gly	Phe	Ile	Val	Gly	Phe			
	195						200					205						
Thr	Arg	Gly	Trp	Lys	Leu	Thr	Leu	Val	Ile	Leu	Ala	Ile	Ser	Pro	Val			
	210					215					220							
Leu	Gly	Leu	Ser	Ala	Ala	Val	Trp	Ala	Lys	Ile	Leu	Ser	Ser	Phe	Thr			
225				230						235					240			
Asp	Lys	Glu	Leu	Leu	Ala	Tyr	Ala	Lys	Ala	Gly	Ala	Val	Ala	Glu	Glu			
			245					250					255					
Val	Leu	Ala	Ala	Ile	Arg	Thr	Val	Ile	Ala	Phe	Gly	Gly	Gln	Lys	Lys			
		260					265					270						
Glu	Leu	Glu	Arg	Tyr	Asn	Lys	Asn	Leu	Glu	Glu	Ala	Lys	Arg	Ile	Gly			
	275						280					285						
Ile	Lys	Lys	Ala	Ile	Thr	Ala	Asn	Ile	Ser	Ile	Gly	Ala	Ala	Phe	Leu			
	290					295					300							
Leu	Ile	Tyr	Ala	Ser	Tyr	Ala	Leu	Ala	Phe	Trp	Tyr	Gly	Thr	Thr	Leu			
305				310						315					320			
Val	Leu	Ser	Gly	Glu	Tyr	Ser	Ile	Gly	Gln	Val	Leu	Thr	Val	Phe	Phe			
			325					330					335					
Ser	Val	Leu	Ile	Gly	Ala	Phe	Ser	Val	Gly	Gln	Ala	Ser	Pro	Ser	Ile			
		340					345					350						
Glu	Ala	Phe	Ala	Asn	Ala	Arg	Gly	Ala	Ala	Tyr	Glu	Ile	Phe	Lys	Ile			
	355						360					365						
Ile	Asp	Asn	Lys	Pro	Ser	Ile	Asp	Ser	Tyr	Ser	Lys	Ser	Gly	His	Lys			
	370					375					380							
Pro	Asp	Asn	Ile	Lys	Gly	Asn	Leu	Glu	Phe	Arg	Asn	Val	His	Phe	Ser			
385				390						395					400			
Tyr	Pro	Ser	Arg	Lys	Glu	Val	Lys	Ile	Leu	Lys	Gly	Leu	Asn	Leu	Lys			
			405					410					415					
Val	Gln	Ser	Gly	Gln	Thr	Val	Ala	Leu	Val	Gly	Asn	Ser	Gly	Cys	Gly			
		420						425				430						
Lys	Ser	Thr	Val	Gln	Leu	Met	Gln	Arg	Leu	Tyr	Asp	Pro	Thr	Glu				
	435					440					445							
Gly	Met	Val	Ser	Val	Asp	Gly	Gln	Asp	Ile	Arg	Thr	Ile	Asn	Val	Arg			
	450				455						460							
Phe	Leu	Arg	Glu	Ile	Ile	Gly	Val	Val	Ser	Gln	Glu	Pro	Val	Leu	Phe			

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

465					470					475					480
Ala	Thr	Thr	Ile	Ala	Glu	Asn	Ile	Arg	Tyr	Gly	Arg	Glu	Asn	Val	Thr
				485					490					495	
Met	Asp	Glu	Ile	Glu	Lys	Ala	Val	Lys	Glu	Ala	Asn	Ala	Tyr	Asp	Phe
			500					505					510		
Ile	Met	Lys	Leu	Pro	His	Lys	Phe	Asp	Thr	Leu	Val	Gly	Glu	Arg	Gly
		515					520					525			
Ala	Gln	Leu	Ser	Gly	Gly	Gln	Lys	Gln	Arg	Ile	Ala	Ile	Ala	Arg	Ala
		530				535					540				
Leu	Val	Arg	Asn	Pro	Lys	Ile	Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser	Ala
545					550					555					560
Leu	Asp	Thr	Glu	Ser	Glu	Ala	Val	Val	Gln	Val	Ala	Leu	Asp	Lys	Ala
				565					570					575	
Arg	Lys	Gly	Arg	Thr	Thr	Ile	Val	Ile	Ala	His	Arg	Leu	Ser	Thr	Val
			580					585					590		
Arg	Asn	Ala	Asp	Val	Ile	Ala	Gly	Phe	Asp	Asp	Gly	Val	Ile	Val	Glu
		595					600					605			
Lys	Gly	Asn	His	Asp	Glu	Leu	Met	Lys	Glu	Lys	Gly	Ile	Tyr	Phe	Lys
		610				615					620				
Leu	Val	Thr	Met	Gln	Thr	Ala	Gly	Asn	Glu	Val	Glu	Leu	Glu	Asn	Ala
625					630					635					640
Ala	Asp	Glu	Ser	Lys	Ser	Glu	Ile	Asp	Ala	Leu	Glu	Met	Ser	Ser	Asn
				645					650					655	
Asp	Ser	Arg	Ser	Ser	Leu	Ile	Arg	Lys	Arg	Ser	Thr	Arg	Arg	Ser	Val
			660					665					670		
Arg	Gly	Ser	Gln	Ala	Gln	Asp	Arg	Lys	Leu	Ser	Thr	Lys	Glu	Ala	Leu
		675				680						685			
Asp	Glu	Ser	Ile	Pro	Pro	Val	Ser	Phe	Trp	Arg	Ile	Met	Lys	Leu	Asn
					695					700					
Leu	Thr	Glu	Trp	Pro	Tyr	Phe	Val	Val	Gly	Val	Phe	Cys	Ala	Ile	Ile
705					710					715					720
Asn	Gly	Gly	Leu	Gln	Pro	Ala	Phe	Ala	Ile	Ile	Phe	Ser	Lys	Ile	Ile
				725					730					735	
Gly	Val	Phe	Thr	Arg	Ile	Asp	Asp	Pro	Glu	Thr	Lys	Arg	Gln	Asn	Ser
			740					745					750		
Asn	Leu	Phe	Ser	Leu	Leu	Phe	Leu	Ala	Leu	Gly	Ile	Ile	Ser	Phe	Ile
		755					760					765			
Thr	Phe	Phe	Leu	Gln	Gly	Phe	Thr	Phe	Gly	Lys	Ala	Gly	Glu	Ile	Leu
		770				775					780				
Thr	Lys	Arg	Leu	Arg	Tyr	Met	Val	Phe	Arg	Ser	Met	Leu	Arg	Gln	Asp
785					790					795					800
Val	Ser	Trp	Phe	Asp	Asp	Pro	Lys	Asn	Thr	Thr	Gly	Ala	Leu	Thr	Thr
				805					810					815	
Arg	Leu	Ala	Asn	Asp	Ala	Ala	Gln	Val	Lys	Gly	Ala	Ile	Gly	Ser	Arg
			820					825					830		
Leu	Ala	Val	Ile	Thr	Gln	Asn	Ile	Ala	Asn	Leu	Gly	Thr	Gly	Ile	Ile
		835					840					845			
Ile	Ser	Phe	Ile	Tyr	Gly	Trp	Gln	Leu	Thr	Leu	Leu	Leu	Leu	Ala	Ile
		850				855					860				
Val	Pro	Ile	Ile	Ala	Ile	Ala	Gly	Val	Val	Glu	Met	Lys	Met	Leu	Ser
865					870					875					880
Gly	Gln	Ala	Leu	Lys	Asp	Lys	Lys	Glu	Leu	Glu	Gly	Ala	Gly	Lys	Ile
				885				890						895	
Ala	Thr	Glu	Ala	Ile	Glu	Asn	Phe	Arg	Thr	Val	Val	Ser	Leu	Thr	Gln
			900					905						910	
Glu	Gln	Lys	Phe	Glu	His	Met	Tyr	Ala	Gln	Ser	Leu	Gln	Val	Pro	Tyr
		915					920					925			

bioRxiv preprint doi: <https://doi.org/10.1101/2020.07.15.201311>; this version posted July 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Arg Asn Ser Leu Arg Lys Ala His Ile Phe Gly Ile Thr Phe Ser Phe
930 935 940
Thr Gln Ala Met Met Tyr Phe Ser Tyr Ala Gly Cys Phe Arg Phe Gly
945 950 955 960
Ala Tyr Leu Val Ala His Lys Leu Met Ser Phe Glu Asp Val Leu Leu
965 970 975
Val Phe Ser Ala Val Val Phe Gly Ala Met Ala Val Gly Gln Val Ser
980 985 990
Ser Phe Ala Pro Asp Tyr Ala Lys Ala Lys Ile Ser Ala Ala His Ile
995 1000 1005
Ile Met Ile Ile Glu Lys Thr Pro Leu Ile Asp Ser Tyr Ser Thr Glu
1010 1015 1020
Gly Leu Met Pro Asn Thr Leu Glu Gly Asn Val Thr Phe Gly Glu Val
1025 1030 1035 1040
Val Phe Asn Tyr Pro Thr Arg Pro Asp Ile Pro Val Leu Gln Gly Leu
1045 1050 1055
Ser Leu Glu Val Lys Lys Gly Gln Thr Leu Ala Leu Val Gly Ser Ser
1060 1065 1070
Gly Cys Gly Lys Ser Thr Val Val Gln Leu Leu Glu Arg Phe Tyr Asp
1075 1080 1085
Pro Leu Ala Gly Lys Val Leu Leu Asp Gly Lys Glu Ile Lys Arg Leu
1090 1095 1100
Asn Val Gln Trp Leu Arg Ala His Leu Gly Ile Val Ser Gln Glu Pro
1105 1110 1115 1120
Ile Leu Phe Asp Cys Ser Ile Ala Glu Asn Ile Ala Tyr Gly Asp Asn
1125 1130 1135
Ser Arg Val Val Ser Gln Glu Glu Ile Val Arg Ala Ala Lys Glu Ala
1140 1145 1150
Asn Ile His Ala Phe Ile Glu Ser Leu Pro Asn Lys Tyr Ser Thr Lys
1155 1160 1165
Val Gly Asp Lys Gly Thr Gln Leu Ser Gly Gly Gln Lys Gln Arg Ile
1170 1175 1180
Ala Ile Ala Arg Ala Leu Val Arg Gln Pro His Ile Leu Leu Leu Asp
1185 1190 1195 1200
Glu Ala Thr Ser Ala Leu Asp Thr Glu Ser Glu Lys Val Val Gln Glu
1205 1210 1215
Ala Leu Asp Lys Ala Arg Glu Gly Arg Thr Cys Ile Val Ile Ala His
1220 1225 1230
Arg Leu Ser Thr Ile Gln Asn Ala Asp Leu Ile Val Val Phe Gln Asn
1235 1240 1245
Gly Arg Val Lys Glu His Gly Thr His Gln Gln Leu Leu Ala Gln Lys
1250 1255 1260
Gly Ile Tyr Phe Ser Met Val Ser Val Gln Ala Gly Thr Lys Arg Gln
1265 1270 1275 1280

<210> 4
<211> 1175
<212> PRT
<213> Mus musculus

<300>
<308> Genbank P06795
<309> 1998-07-15

<400> 4
Ser Asn Ser Ser Leu Glu Glu Glu Met Ala Ile Tyr Ala Tyr Tyr Tyr
1 5 10 15

Thr Gly Ile Gly Ala Gly Val Leu Ile Val Ala Tyr Ile Gln Val Ser
 20 25 30
 Leu Trp Cys Leu Ala Ala Gly Arg Gln Ile His Lys Ile Arg Gln Lys
 35 40 45
 Phe Phe His Ala Ile Met Asn Gln Glu Ile Gly Trp Phe Asp Val His
 50 55 60
 Asp Val Gly Glu Leu Asn Thr Arg Leu Thr Asp Asp Val Ser Lys Ile
 65 70 75 80
 Asn Asp Gly Ile Gly Asp Lys Ile Gly Met Phe Phe Gln Ser Ile Thr
 85 90 95
 Thr Phe Leu Ala Gly Phe Ile Ile Gly Phe Ile Ser Gly Trp Lys Leu
 100 105 110
 Thr Leu Val Ile Leu Ala Val Ser Pro Leu Ile Gly Leu Ser Ser Ala
 115 120 125
 Leu Trp Ala Lys Val Leu Thr Ser Phe Thr Asn Lys Glu Leu Gln Ala
 130 135 140
 Tyr Ala Lys Ala Gly Ala Val Ala Glu Glu Val Leu Ala Ala Ile Arg
 145 150 155 160
 Thr Val Ile Ala Phe Gly Gly Gln Gln Lys Glu Leu Glu Arg Tyr Asn
 165 170 175
 Lys Asn Leu Glu Glu Ala Lys Asn Val Gly Ile Lys Lys Ala Ile Thr
 180 185 190
 Ala Ser Ile Ser Ile Gly Ile Ala Tyr Leu Leu Val Tyr Ala Ser Tyr
 195 200 205
 Ala Leu Ala Phe Trp Tyr Gly Thr Ser Leu Val Leu Ser Asn Glu Tyr
 210 215 220
 Ser Ile Gly Glu Val Leu Thr Val Phe Phe Ser Ile Leu Leu Gly Thr
 225 230 235 240
 Phe Ser Ile Gly His Leu Ala Pro Asn Ile Glu Ala Phe Ala Asn Ala
 245 250 255
 Arg Gly Ala Ala Phe Glu Ile Phe Lys Ile Ile Asp Asn Glu Pro Ser
 260 265 270
 Ile Asp Ser Phe Ser Thr Lys Gly Tyr Lys Pro Asp Ser Ile Met Gly
 275 280 285
 Asn Leu Glu Phe Lys Asn Val His Phe Asn Tyr Pro Ser Arg Ser Glu
 290 295 300
 Val Gln Ile Leu Lys Gly Leu Asn Leu Lys Val Lys Ser Gly Gln Thr
 305 310 315 320
 Val Ala Leu Val Gly Asn Ser Gly Cys Gly Lys Ser Thr Thr Val Gln
 325 330 335
 Leu Met Gln Arg Leu Tyr Asp Pro Leu Glu Gly Val Val Ser Ile Asp
 340 345 350
 Gly Gln Asp Ile Arg Thr Ile Asn Val Arg Tyr Leu Arg Glu Ile Ile
 355 360 365
 Gly Val Val Ser Gln Glu Pro Val Leu Phe Ala Thr Thr Ile Ala Glu
 370 375 380
 Asn Ile Arg Tyr Gly Arg Glu Asp Val Thr Met Asp Glu Ile Glu Lys
 385 390 395 400
 Ala Val Lys Glu Ala Asn Ala Tyr Asp Phe Ile Met Lys Leu Pro His
 405 410 415
 Gln Phe Asp Thr Leu Val Gly Glu Arg Gly Ala Gln Leu Ser Gly Gly
 420 425 430
 Gln Lys Gln Arg Ile Ala Ile Ala Arg Ala Leu Val Arg Asn Pro Lys
 435 440 445
 Ile Leu Leu Leu Asp Glu Ala Thr Ser Ala Leu Asp Thr Glu Ser Glu
 450 455 460
 Ala Val Val Gln Ala Ala Leu Asp Lys Ala Arg Glu Gly Arg Thr Thr

465					470					475					480
Ile	Val	Ile	Ala	His	Arg	Leu	Ser	Thr	Val	Arg	Asn	Ala	Asp	Val	Ile
				485						490				495	
Ala	Gly	Phe	Asp	Gly	Gly	Val	Ile	Val	Glu	Gln	Gly	Asn	His	Asp	Glu
			500					505					510		
Leu	Met	Arg	Glu	Lys	Gly	Ile	Tyr	Phe	Lys	Leu	Val	Met	Thr	Gln	Thr
		515					520					525			
Arg	Gly	Asn	Glu	Ile	Glu	Pro	Gly	Asn	Asn	Ala	Tyr	Gly	Ser	Gln	Ser
		530				535					540				
Asp	Thr	Asp	Ala	Ser	Glu	Leu	Thr	Ser	Glu	Glu	Ser	Lys	Ser	Pro	Leu
545					550					555					560
Ile	Arg	Arg	Ser	Ile	Tyr	Arg	Ser	Val	His	Arg	Lys	Gln	Asp	Gln	Glu
			565						570					575	
Arg	Arg	Leu	Ser	Met	Lys	Glu	Ala	Val	Asp	Glu	Asp	Val	Pro	Leu	Val
		580						585					590		
Ser	Phe	Trp	Arg	Ile	Leu	Asn	Leu	Asn	Leu	Ser	Glu	Trp	Pro	Tyr	Leu
		595				600						605			
Leu	Val	Gly	Val	Leu	Cys	Ala	Val	Ile	Asn	Gly	Cys	Ile	Gln	Pro	Val
	610					615					620				
Phe	Ala	Ile	Val	Phe	Ser	Arg	Ile	Val	Gly	Val	Phe	Ser	Arg	Asp	Asp
625					630					635					640
Asp	His	Glu	Thr	Lys	Arg	Gln	Asn	Cys	Asn	Leu	Phe	Ser	Leu	Phe	Phe
				645					650					655	
Leu	Val	Met	Gly	Leu	Ile	Ser	Phe	Val	Thr	Tyr	Phe	Phe	Gln	Gly	Phe
		660						665					670		
Thr	Phe	Gly	Lys	Ala	Gly	Glu	Ile	Leu	Thr	Lys	Arg	Val	Arg	Tyr	Met
		675				680						685			
Val	Phe	Lys	Ser	Met	Leu	Arg	Gln	Asp	Ile	Ser	Trp	Phe	Asp	Asp	His
	690					695					700				
Lys	Asn	Ser	Thr	Gly	Ser	Leu	Thr	Thr	Arg	Leu	Ala	Ser	Asp	Ala	Ser
705					710					715					720
Ser	Val	Lys	Gly	Ala	Met	Gly	Ala	Arg	Leu	Ala	Val	Val	Thr	Gln	Asn
				725					730					735	
Val	Ala	Asn	Leu	Gly	Thr	Gly	Val	Ile	Leu	Ser	Leu	Val	Tyr	Gly	Trp
		740						745					750		
Gln	Leu	Thr	Leu	Leu	Leu	Val	Val	Ile	Ile	Pro	Leu	Ile	Val	Leu	Gly
		755				760						765			
Gly	Ile	Ile	Glu	Met	Lys	Leu	Leu	Ser	Gly	Gln	Ala	Leu	Lys	Asp	Lys
	770					775					780				
Lys	Gln	Leu	Glu	Ile	Ser	Gly	Lys	Ile	Ala	Thr	Glu	Ala	Ile	Glu	Asn
785					790					795					800
Phe	Arg	Thr	Ile	Val	Ser	Leu	Thr	Arg	Glu	Gln	Lys	Phe	Glu	Thr	Met
			805						810					815	
Tyr	Ala	Gln	Ser	Leu	Gln	Val	Pro	Tyr	Arg	Asn	Ala	Met	Lys	Lys	Ala
			820					825					830		
His	Val	Phe	Gly	Ile	Thr	Phe	Ser	Phe	Thr	Gln	Ala	Met	Met	Tyr	Phe
		835				840						845			
Ser	Tyr	Ala	Ala	Cys	Phe	Arg	Phe	Gly	Ala	Tyr	Leu	Val	Ala	Gln	Gln
	850					855					860				
Leu	Met	Thr	Phe	Glu	Asn	Val	Met	Leu	Val	Phe	Ser	Ala	Val	Val	Phe
865					870					875					880
Gly	Ala	Met	Ala	Ala	Gly	Asn	Thr	Ser	Ser	Phe	Ala	Pro	Asp	Tyr	Ala
				885					890					895	
Lys	Ala	Lys	Val	Ser	Ala	Ser	His	Ile	Arg	Ile	Ile	Glu	Lys	Thr	
		900						905				910			
Pro	Glu	Ile	Asp	Ser	Tyr	Ser	Thr	Glu	Gly	Leu	Lys	Pro	Thr	Leu	Leu
		915					920					925			

Glu Gly Asn Val Lys Phe Asn Gly Val Gln Phe Asn Tyr Pro Thr Arg
 930 935 940
 Pro Asn Ile Pro Val Leu Gln Gly Leu Ser Leu Glu Val Lys Lys Gly
 945 950 955 960
 Gln Thr Leu Ala Leu Val Gly Ser Ser Gly Cys Gly Lys Ser Thr Val
 965 970 975
 Val Gln Leu Leu Glu Arg Phe Tyr Asp Pro Met Ala Gly Ser Val Phe
 980 985 990
 Leu Asp Gly Lys Glu Ile Lys Gln Leu Asn Val Gln Trp Leu Arg Ala
 995 1000 1005
 His Leu Gly Ile Val Ser Gln Glu Pro Ile Leu Phe Asp Cys Ser Ile
 1010 1015 1020
 Ala Glu Asn Ile Ala Tyr Gly Asp Asn Ser Arg Ala Val Ser His Glu
 1025 1030 1035 1040
 Glu Ile Val Arg Ala Ala Lys Glu Ala Asn Ile His Gln Phe Ile Asp
 1045 1050 1055
 Ser Leu Pro Asp Lys Tyr Asn Thr Arg Val Gly Asp Lys Gly Thr Gln
 1060 1065 1070
 Leu Ser Gly Gly Gln Lys Gln Arg Ile Ala Ile Ala Arg Ala Leu Val
 1075 1080 1085
 Arg Gln Pro His Ile Leu Leu Leu Asp Glu Ala Thr Ser Ala Leu Asp
 1090 1095 1100
 Thr Glu Ser Glu Lys Val Val Gln Glu Ala Leu Asp Lys Ala Arg Glu
 1105 1110 1115 1120
 Gly Arg Thr Cys Ile Val Ile Ala His Arg Leu Ser Thr Ile Gln Asn
 1125 1130 1135
 Ala Asp Leu Ile Val Val Ile Glu Asn Gly Lys Val Lys Glu His Gly
 1140 1145 1150
 Thr His Gln Gln Leu Leu Ala Gln Lys Gly Ile Tyr Phe Ser Met Val
 1155 1160 1165
 Gln Ala Gly Ala Lys Arg Ser
 1170 1175

<210> 5

<211> 1066

<212> PRT

<213> Homo sapiens

<300>

<308> Genbank P21439

<309> 1998-07-15

<400> 5

Trp Lys Leu Thr Leu Val Ile Met Ala Ile Ser Pro Ile Leu Gly Leu
 1 5 10 15
 Ser Ala Ala Val Trp Ala Lys Ile Leu Ser Ala Phe Ser Asp Lys Glu
 20 25 30
 Leu Ala Ala Tyr Ala Lys Ala Gly Ala Val Ala Glu Glu Ala Leu Gly
 35 40 45
 Ala Ile Arg Thr Val Ile Ala Phe Gly Gly Gln Asn Lys Glu Leu Glu
 50 55 60
 Arg Tyr Gln Lys His Leu Glu Asn Ala Lys Glu Ile Gly Ile Lys Lys
 65 70 75 80
 Ala Ile Ser Ala Asn Ile Ser Met Gly Ile Ala Phe Leu Leu Ile Tyr
 85 90 95
 Ala Ser Tyr Ala Leu Ala Phe Trp Tyr Gly Ser Thr Leu Val Ile Ser
 100 105 110

Lys Glu Tyr Thr Ile Gly Asn Ala Met Thr Val Phe Phe Ser Ile Leu
 115 120 125
 Ile Gly Ala Phe Ser Val Gly Gln Ala Ala Pro Cys Ile Asp Ala Phe
 130 135 140
 Ala Asn Ala Arg Gly Ala Ala Tyr Val Ile Phe Asp Ile Ile Asp Asn
 145 150 155 160
 Asn Pro Lys Ile Asp Ser Phe Ser Glu Arg Gly His Lys Pro Asp Ser
 165 170 175
 Ile Lys Gly Asn Leu Glu Phe Asn Asp Val His Phe Ser Tyr Pro Ser
 180 185 190
 Arg Ala Asn Val Lys Ile Leu Lys Gly Leu Asn Leu Lys Val Gln Ser
 195 200 205
 Gly Gln Thr Val Ala Leu Val Gly Ser Ser Gly Cys Gly Lys Ser Thr
 210 215 220
 Thr Val Gln Leu Ile Gln Arg Leu Tyr Asp Pro Asp Glu Gly Thr Ile
 225 230 235 240
 Asn Ile Asp Gly Gln Asp Ile Arg Asn Phe Asn Val Asn Tyr Leu Arg
 245 250 255
 Glu Ile Ile Gly Val Val Ser Gln Glu Pro Val Leu Phe Ser Thr Thr
 260 265 270
 Ile Ala Glu Asn Ile Cys Tyr Gly Arg Gly Asn Val Thr Met Asp Glu
 275 280 285
 Ile Lys Lys Ala Val Lys Glu Ala Asn Ala Tyr Glu Phe Ile Met Lys
 290 295 300
 Leu Pro Gln Lys Phe Asp Thr Leu Val Gly Glu Arg Gly Ala Gln Leu
 305 310 315 320
 Ser Gly Gly Gln Lys Gln Arg Ile Ala Ile Ala Arg Ala Leu Val Arg
 325 330 335
 Asn Pro Lys Ile Leu Leu Leu Asp Glu Ala Thr Ser Ala Leu Asp Thr
 340 345 350
 Glu Ser Glu Ala Glu Val Gln Ala Ala Leu Asp Lys Ala Arg Glu Gly
 355 360 365
 Arg Thr Thr Ile Val Ile Ala His Arg Leu Ser Thr Val Arg Asn Ala
 370 375 380
 Asp Val Ile Ala Gly Phe Glu Asp Gly Val Ile Val Glu Gln Gly Ser
 385 390 395 400
 His Ser Glu Leu Met Lys Lys Glu Gly Val Tyr Phe Lys Leu Val Asn
 405 410 415
 Met Gln Thr Ser Gly Ser Gln Ile Gln Ser Glu Glu Phe Glu Leu Asn
 420 425 430
 Asp Glu Lys Ala Ala Thr Arg Met Ala Pro Asn Gly Trp Lys Ser Arg
 435 440 445
 Leu Phe Arg His Ser Thr Gln Lys Asn Leu Lys Asn Ser Gln Met Cys
 450 455 460
 Gln Lys Ser Leu Asp Val Glu Thr Asp Gly Leu Glu Ala Asn Val Pro
 465 470 475 480
 Pro Val Ser Phe Leu Lys Val Leu Lys Leu Asn Lys Thr Glu Trp Pro
 485 490 495
 Tyr Phe Val Val Gly Thr Val Cys Ala Ile Ala Asn Gly Gly Leu Gln
 500 505 510
 Pro Ala Phe Ser Val Ile Phe Ser Glu Ile Ile Ala Ile Phe Gly Pro
 515 520 525
 Gly Asp Asp Ala Val Lys Gln Gln Lys Cys Asn Ile Phe Ser Leu Ile
 530 535 540
 Phe Leu Phe Leu Gly Ile Ile Ser Phe Phe Thr Phe Phe Leu Gln Gly
 545 550 555 560
 Phe Thr Phe Gly Lys Ala Gly Glu Ile Leu Thr Arg Arg Leu Arg Ser

Met	Ala	Phe	Lys	Ala	Met	Leu	Arg	Gln	Asp	Met	Ser	Trp	Phe	Asp	Asp
			580					585					590		
His	Lys	Asn	Ser	Thr	Gly	Ala	Leu	Ser	Thr	Arg	Leu	Ala	Thr	Asp	Ala
		595					600					605			
Ala	Gln	Val	Gln	Gly	Ala	Thr	Gly	Thr	Arg	Leu	Ala	Leu	Ile	Ala	Gln
	610				615					620					
Asn	Ile	Ala	Asn	Leu	Gly	Thr	Gly	Ile	Ile	Ile	Ser	Phe	Ile	Tyr	Gly
625					630					635					640
Trp	Gln	Leu	Thr	Leu	Leu	Leu	Leu	Ala	Val	Val	Pro	Ile	Ile	Ala	Val
			645						650					655	
Ser	Gly	Ile	Val	Glu	Met	Lys	Leu	Leu	Ala	Gly	Asn	Ala	Lys	Arg	Asp
			660					665					670		
Lys	Lys	Glu	Leu	Glu	Ala	Ala	Gly	Lys	Ile	Ala	Thr	Glu	Ala	Ile	Glu
		675					680					685			
Asn	Ile	Arg	Thr	Val	Val	Ser	Leu	Thr	Gln	Glu	Arg	Lys	Phe	Glu	Ser
	690					695					700				
Met	Tyr	Val	Glu	Lys	Leu	Tyr	Gly	Pro	Tyr	Arg	Asn	Ser	Val	Gln	Lys
705					710					715					720
Ala	His	Ile	Tyr	Gly	Ile	Thr	Phe	Ser	Ile	Ser	Gln	Ala	Phe	Met	Tyr
			725						730					735	
Phe	Ser	Tyr	Ala	Gly	Cys	Phe	Arg	Phe	Gly	Ala	Tyr	Leu	Ile	Val	Asn
			740					745					750		
Gly	His	Met	Arg	Phe	Arg	Asp	Val	Ile	Leu	Val	Phe	Ser	Ala	Ile	Val
		755					760				765				
Phe	Gly	Ala	Val	Ala	Leu	Gly	His	Ala	Ser	Ser	Phe	Ala	Pro	Asp	Tyr
	770					775					780				
Ala	Lys	Ala	Lys	Leu	Ser	Ala	Ala	His	Leu	Phe	Met	Leu	Phe	Glu	Arg
785					790					795					800
Gln	Pro	Leu	Ile	Asp	Ser	Tyr	Ser	Glu	Glu	Gly	Leu	Lys	Pro	Asp	Lys
				805					810					815	
Phe	Glu	Gly	Asn	Ile	Thr	Phe	Asn	Glu	Val	Val	Phe	Asn	Tyr	Pro	Thr
			820					825					830		
Arg	Ala	Asn	Val	Pro	Val	Leu	Gln	Gly	Leu	Ser	Leu	Glu	Val	Lys	Lys
		835					840					845			
Gly	Gln	Thr	Leu	Ala	Leu	Val	Gly	Ser	Ser	Gly	Cys	Gly	Lys	Ser	Thr
	850					855					860				
Val	Val	Gln	Leu	Leu	Glu	Arg	Phe	Tyr	Asp	Pro	Leu	Ala	Gly	Thr	Val
865					870					875					880
Leu	Leu	Asp	Gly	Gln	Glu	Ala	Lys	Lys	Leu	Asn	Val	Gln	Trp	Leu	Arg
				885					890					895	
Ala	Gln	Leu	Gly	Ile	Val	Ser	Gln	Glu	Pro	Ile	Leu	Phe	Asp	Cys	Ser
			900					905					910		
Ile	Ala	Glu	Asn	Ile	Ala	Tyr	Gly	Asp	Asn	Ser	Arg	Val	Val	Ser	Gln
		915					920					925			
Asp	Glu	Ile	Val	Ser	Ala	Ala	Lys	Ala	Ala	Asn	Ile	His	Pro	Phe	Ile
	930					935					940				
Glu	Thr	Leu	Pro	His	Lys	Tyr	Glu	Thr	Arg	Val	Gly	Asp	Lys	Gly	Thr
945					950					955					960
Gln	Leu	Ser	Gly	Gly	Gln	Lys	Gln	Arg	Ile	Ala	Ile	Ala	Arg	Ala	Leu
				965					970					975	
Ile	Arg	Gln	Pro	Gln	Ile	Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser	Ala	Leu
			980					985					990		
Asp	Thr	Glu	Ser	Glu	Lys	Val	Val	Gln	Glu	Ala	Leu	Asp	Lys	Ala	Arg
		995					1000					1005			
Glu	Gly	Arg	Thr	Cys	Ile	Val	Ile	Ala	His	Arg	Leu	Ser	Thr	Ile	Gln
	1010					1015						1020			

Asn Ala Asp Leu Ile Val Val Phe Gln Asn Gly Arg Val Lys Glu His
 1025 1030 1035 1040
 Gly Thr His Gln Gln Leu Leu Ala Gln Lys Gly Ile Tyr Phe Ser Met
 1045 1050 1055
 Val Ser Val Gln Ala Gly Thr Gln Asn Leu
 1060 1065

<210> 6
 <211> 1266
 <212> PRT
 <213> Mus musculus

<300>
 <308> Genbank P21440
 <309> 1997-11-01

<400> 6

Trp Lys Leu Thr Leu Val Ile Met Ala Ile Ser Pro Ile Leu Gly Leu
 1 5 10 15
 Ser Thr Ala Val Trp Ala Lys Ile Leu Ser Thr Phe Ser Asp Lys Glu
 20 25 30
 Leu Ala Ala Tyr Ala Lys Ala Gly Ala Val Ala Glu Glu Ala Pro Gly
 35 40 45
 Ala Ile Arg Thr Val Ile Ala Phe Gly Gly Gln Asn Lys Glu Leu Glu
 50 55 60
 Arg Tyr Gln Lys His Leu Glu Asn Ala Lys Lys Ile Gly Ile Lys Lys
 65 70 75 80
 Ala Ile Ser Ala Asn Ile Ser Met Gly Ile Ala Phe Leu Leu Ile Tyr
 85 90 95
 Ala Ser Tyr Ala Leu Ala Phe Trp Tyr Gly Ser Thr Leu Val Ile Ser
 100 105 110
 Lys Glu Tyr Thr Ile Gly Asn Ala Met Thr Val Phe Phe Ser Ile Leu
 115 120 125
 Ile Gly Ala Phe Ser Val Gly Gln Ala Ala Pro Cys Ile Asp Ala Phe
 130 135 140
 Ala Asn Ala Arg Gly Ala Ala Tyr Val Ile Phe Asp Ile Ile Asp Asn
 145 150 155 160
 Asn Pro Lys Ile Asp Ser Phe Ser Glu Arg Gly His Lys Pro Asp Asn
 165 170 175
 Ile Lys Gly Asn Leu Glu Phe Ser Asp Val His Phe Ser Tyr Pro Ser
 180 185 190
 Arg Ala Asn Ile Lys Ile Leu Lys Gly Leu Asn Leu Lys Val Lys Ser
 195 200 205
 Gly Gln Thr Val Ala Leu Val Gly Asn Ser Gly Cys Gly Lys Ser Thr
 210 215 220
 Thr Val Gln Leu Leu Gln Arg Leu Tyr Asp Pro Thr Glu Gly Lys Ile
 225 230 235 240
 Ser Ile Asp Gly Gln Asp Ile Arg Asn Phe Asn Val Arg Cys Leu Arg
 245 250 255
 Glu Ile Ile Gly Val Val Ser Gln Glu Pro Val Leu Phe Ser Thr Thr
 260 265 270
 Ile Ala Glu Asn Ile Arg Tyr Gly Arg Gly Asn Val Thr Met Asp Glu
 275 280 285
 Ile Glu Lys Ala Val Lys Glu Ala Asn Ala Tyr Asp Phe Ile Met Lys
 290 295 300
 Leu Pro Gln Lys Phe Asp Thr Leu Val Gly Asp Arg Gly Ala Gln Leu
 305 310 315 320

bioRxiv preprint doi: <https://doi.org/10.1101/1266>; this version posted November 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ser	Gly	Gly	Gln	Lys	Gln	Arg	Ile	Ala	Ile	Ala	Arg	Ala	Leu	Val	Arg		
				325					330					335			
Asn	Pro	Lys	Ile	Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser	Ala	Leu	Asp	Thr		
			340					345					350				
Glu	Ser	Glu	Ala	Glu	Val	Gln	Ala	Ala	Leu	Asp	Lys	Ala	Arg	Glu	Gly		
			355				360					365					
Arg	Thr	Thr	Ile	Val	Ile	Ala	His	Arg	Leu	Ser	Thr	Ile	Arg	Asn	Ala		
			370			375					380						
Asp	Val	Ile	Ala	Gly	Phe	Glu	Asp	Gly	Val	Ile	Val	Glu	Gln	Gly	Ser		
385					390					395					400		
His	Ser	Glu	Leu	Met	Lys	Lys	Glu	Gly	Ile	Tyr	Phe	Arg	Leu	Val	Asn		
				405					410					415			
Met	Gln	Thr	Ala	Gly	Ser	Gln	Ile	Leu	Ser	Glu	Glu	Phe	Glu	Ala	Arg		
			420					425					430				
Ala	Leu	Val	Arg	Asn	Pro	Lys	Ile	Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser		
			435			440						445					
Ala	Leu	Asp	Thr	Glu	Ser	Glu	Ala	Val	Val	Gln	Val	Ala	Leu	Asp	Lys		
			450			455					460						
Ala	Arg	Lys	Gly	Arg	Thr	Thr	Ile	Val	Ile	Ala	His	Arg	Leu	Ser	Thr		
465					470					475					480		
Val	Arg	Asn	Ala	Asp	Val	Ile	Ala	Gly	Phe	Asp	Asp	Gly	Val	Ile	Val		
				485					490					495			
Glu	Lys	Gly	Asn	His	Asp	Glu	Leu	Met	Lys	Glu	Lys	Gly	Ile	Tyr	Phe		
			500					505					510				
Lys	Leu	Val	Thr	Met	Gln	Thr	Ala	Gly	Asn	Glu	Val	Glu	Leu	Glu	Asn		
			515				520					525					
Ala	Ala	Ala	Arg	Ala	Leu	Val	Arg	Asn	Pro	Lys	Ile	Leu	Leu	Leu	Asp		
			530			535				540							
Glu	Ala	Thr	Ser	Ala	Leu	Asp	Thr	Glu	Ser	Glu	Ala	Val	Val	Gln	Ala		
545				550						555					560		
Ala	Leu	Asp	Lys	Ala	Arg	Glu	Gly	Arg	Thr	Thr	Ile	Val	Ile	Ala	His		
				565					570					575			
Arg	Leu	Ser	Thr	Val	Arg	Asn	Ala	Asp	Val	Ile	Ala	Gly	Phe	Asp	Gly		
			580					585					590				
Gly	Val	Ile	Val	Glu	Gln	Gly	Asn	His	Asp	Glu	Leu	Met	Arg	Glu	Lys		
			595				600						605				
Gly	Ile	Tyr	Phe	Lys	Leu	Val	Met	Thr	Gln	Thr	Arg	Gly	Asn	Glu	Ile		
			610			615					620						
Glu	Pro	Gly	Asn	Asn	Ala	Val	Glu	Leu	Ser	Asp	Glu	Lys	Ala	Ala	Gly		
625				630						635					640		
Asp	Val	Ala	Pro	Asn	Gly	Trp	Lys	Ala	Arg	Ile	Phe	Arg	Asn	Ser	Thr		
				645					650					655			
Lys	Lys	Ser	Leu	Lys	Ser	Pro	His	Gln	Asn	Arg	Leu	Asp	Glu	Glu	Thr		
			660				665						670				
Asn	Glu	Leu	Asp	Ala	Asn	Val	Pro	Pro	Val	Ser	Phe	Leu	Lys	Val	Leu		
			675			680						685					
Lys	Leu	Asn	Lys	Thr	Glu	Trp	Pro	Tyr	Phe	Val	Val	Gly	Thr	Val	Cys		
			690			695						700					
Ala	Ile	Ala	Asn	Gly	Ala	Leu	Gln	Pro	Ala	Phe	Ser	Ile	Ile	Leu	Ser		
705				710						715					720		
Glu	Met	Ile	Ala	Ile	Phe	Gly	Pro	Gly	Asp	Asp	Ala	Val	Lys	Gln	Gln		
				725					730					735			
Lys	Cys	Asn	Met	Phe	Ser	Leu	Val	Phe	Leu	Gly	Leu	Gly	Val	Leu	Ser		
			740					745					750				
Phe	Phe	Thr	Phe	Phe	Leu	Gln	Gly	Phe	Thr	Phe	Gly	Lys	Ala	Gly	Glu		
			755			760						765					
Ile	Leu	Thr	Thr	Arg	Leu	Arg	Ser	Met	Ala	Phe	Lys	Ala	Met	Leu	Arg		

770	775	780
Gln Asp Met Ser Trp Phe	Asp Asp His Lys Asn Ser Thr Gly Ala Leu	
785	790	795
Ser Thr Arg Leu Ala Thr	Asp Ala Ala Gln Val Gln Gly Ala Thr Gly	800
	805	810
Thr Lys Leu Ala Leu Ile	Ala Gln Asn Thr Ala Asn Leu Gly Thr Gly	815
	820	825
Ile Ile Ile Ser Phe Ile	Tyr Gly Trp Gln Leu Thr Leu Leu Leu Leu	830
	835	840
Ser Val Val Pro Phe Ile	Ala Val Ala Gly Ile Val Glu Met Lys Met	845
	850	855
Leu Ala Gly Asn Ala Lys	Arg Asp Lys Lys Glu Met Glu Ala Ala Gly	860
865	870	875
Lys Ile Ala Thr Glu Ala	Ile Glu Asn Ile Arg Thr Val Val Ser Leu	880
	885	890
Thr Gln Glu Arg Lys Phe	Glu Ser Met Tyr Val Glu Lys Leu His Gly	895
	900	905
Pro Tyr Arg Asn Ser Val	Arg Lys Ala His Ile Tyr Gly Ile Thr Phe	910
	915	920
Ser Ile Ser Gln Ala Phe	Met Tyr Phe Ser Tyr Ala Gly Cys Phe Arg	925
	930	935
Phe Gly Ser Tyr Leu Ile	Val Asn Gly His Met Arg Phe Lys Asp Val	940
945	950	955
Ile Leu Val Phe Ser Ala	Ile Val Leu Gly Ala Val Ala Leu Gly His	960
	965	970
Ala Ser Ser Phe Ala Pro	Asp Tyr Ala Lys Ala Lys Leu Ser Ala Ala	975
	980	985
Tyr Leu Phe Ser Leu Phe	Glu Arg Gln Pro Leu Ile Asp Ser Tyr Ser	990
	995	1000
Gly Glu Gly Leu Trp Pro	Asp Lys Phe Glu Gly Ser Val Thr Phe Asn	1005
	1010	1015
Glu Val Val Phe Asn Tyr	Pro Thr Arg Ala Asn Val Pro Val Leu Gln	1020
1025	1030	1035
Gly Leu Ser Leu Glu Val	Lys Lys Gly Gln Thr Leu Ala Leu Val Gly	1040
	1045	1050
Ser Ser Gly Cys Gly Lys	Ser Thr Val Val Gln Leu Leu Glu Arg Phe	1055
	1060	1065
Tyr Asp Pro Met Ala Gly	Ser Val Leu Leu Asp Gly Gln Glu Ala Lys	1070
	1075	1080
Lys Leu Asn Val Gln Trp	Leu Arg Ala Gln Leu Gly Ile Val Ser Gln	1085
	1090	1095
Glu Pro Ile Leu Phe Asp	Cys Ser Ile Ala Glu Asn Ile Ala Tyr Gly	1100
1105	1110	1115
Asp Asn Ser Arg Val Val	Pro His Asp Glu Ile Val Arg Ala Ala Lys	1120
	1125	1130
Glu Ala Asn Ile His Pro	Phe Ile Glu Thr Leu Pro Gln Lys Tyr Asn	1135
	1140	1145
Thr Arg Val Gly Asp Lys	Gly Thr Gln Leu Ser Gly Gly Gln Lys Gln	1150
	1155	1160
Arg Ile Ala Ile Ala Arg	Ala Leu Ile Arg Gln Pro Arg Val Leu Leu	1165
	1170	1175
Leu Asp Glu Ala Thr Ser	Ala Leu Asp Thr Glu Ser Glu Lys Val Val	1180
1185	1190	1195
Gln Glu Ala Leu Asp Lys	Ala Arg Glu Gly Arg Thr Cys Ile Val Ile	1200
	1205	1210
Ala His Arg Leu Ser Thr	Ile Gln Asn Ala Asp Leu Ile Val Val Ile	1215
	1220	1225
		1230

Glu Asn Gly Lys Val Lys Glu His Gly Thr His Gln Gln Leu Leu Ala
 1235 1240 1245
 Gln Lys Gly Ile Tyr Phe Ser Met Val Asn Ile Gln Ala Gly Thr Gln
 1250 1255 1260
 Asn Leu
 1265

<210> 7
 <211> 1207
 <212> PRT
 <213> Arabidopsis thaliana

<300>
 <308> Genbank A42150
 <309> 1997-03-13

<400> 7

Glu	Lys	Met	Met	Glu	Glu	Val	Leu	Lys	Tyr	Ala	Leu	Tyr	Phe	Leu	Val
1				5					10					15	
Val	Gly	Ala	Ala	Ile	Trp	Ala	Ser	Ser	Trp	Ala	Glu	Ile	Ser	Cys	Trp
			20					25					30		
Met	Trp	Ser	Gly	Glu	Arg	Gln	Thr	Thr	Lys	Met	Arg	Ile	Lys	Tyr	Leu
			35				40					45			
Glu	Ala	Ala	Leu	Asn	Gln	Asp	Ile	Gln	Phe	Phe	Asp	Thr	Glu	Val	Arg
	50				55				60						
Thr	Ser	Asp	Val	Val	Phe	Ala	Ile	Asn	Thr	Asp	Ala	Val	Met	Val	Gln
65					70				75						80
Asp	Ala	Ile	Ser	Glu	Lys	Leu	Gly	Asn	Phe	Ile	His	Tyr	Met	Ala	Thr
				85					90					95	
Phe	Val	Ser	Gly	Phe	Ile	Val	Gly	Phe	Thr	Ala	Val	Trp	Gln	Leu	Ala
			100					105					110		
Leu	Val	Thr	Leu	Ala	Val	Val	Pro	Leu	Ile	Ala	Val	Ile	Gly	Gly	Ile
			115				120						125		
His	Thr	Thr	Thr	Leu	Ser	Lys	Leu	Ser	Asn	Lys	Ser	Gln	Glu	Ser	Leu
	130					135					140				
Ser	Gln	Ala	Gly	Asn	Ile	Val	Glu	Gln	Thr	Val	Val	Gln	Ile	Arg	Val
145					150					155					160
Val	Met	Ala	Phe	Val	Gly	Glu	Ser	Arg	Ala	Ser	Gln	Ala	Tyr	Ser	Ser
				165					170					175	
Ala	Leu	Lys	Ile	Ala	Gln	Lys	Leu	Gly	Tyr	Lys	Thr	Gly	Leu	Ala	Lys
			180					185					190		
Gly	Met	Gly	Leu	Gly	Ala	Thr	Tyr	Phe	Val	Val	Phe	Cys	Cys	Tyr	Ala
	195					200					205				
Leu	Leu	Leu	Trp	Tyr	Gly	Gly	Tyr	Leu	Val	Arg	His	His	Leu	Thr	Asn
	210					215					220				
Gly	Gly	Leu	Ala	Ile	Ala	Thr	Met	Phe	Ala	Val	Met	Ile	Gly	Gly	Leu
225					230					235					240
Ala	Leu	Gly	Gln	Ser	Ala	Pro	Ser	Met	Ala	Ala	Phe	Ala	Lys	Ala	Lys
				245					250					255	
Val	Ala	Ala	Ala	Lys	Ile	Phe	Arg	Ile	Ile	Asp	His	Lys	Pro	Thr	Ile
			260				265						270		
Glu	Arg	Asn	Ser	Glu	Ser	Gly	Val	Glu	Leu	Asp	Ser	Val	Thr	Gly	Leu
	275					280						285			
Val	Glu	Leu	Lys	Asn	Val	Asp	Phe	Ser	Tyr	Pro	Ser	Arg	Pro	Asp	Val
	290					295					300				
Lys	Ile	Leu	Asn	Asn	Phe	Cys	Leu	Ser	Val	Pro	Ala	Gly	Lys	Thr	Ile
305					310					315					320

bioRxiv preprint doi: <https://doi.org/10.1101/1207>; this version posted March 13, 1997. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ala	Leu	Val	Gly	Ser	Ser	Gly	Ser	Gly	Lys	Ser	Thr	Val	Val	Ser	Leu			
				325					330					335				
Ile	Glu	Arg	Phe	Tyr	Asp	Pro	Asn	Ser	Gly	Gln	Val	Leu	Leu	Asp	Gly			
			340					345					350					
Gln	Asp	Leu	Lys	Thr	Leu	Lys	Leu	Arg	Trp	Leu	Arg	Gln	Gln	Ile	Gly			
		355					360					365						
Leu	Val	Ser	Gln	Glu	Pro	Ala	Leu	Phe	Ala	Thr	Ser	Ile	Lys	Glu	Asn			
	370					375					380							
Ile	Leu	Leu	Gly	Arg	Pro	Asp	Ala	Asp	Gln	Val	Glu	Ile	Glu	Glu	Ala			
385					390					395					400			
Ala	Arg	Val	Ala	Asn	Ala	His	Ser	Phe	Ile	Ile	Lys	Leu	Pro	Asp	Gly			
			405					410						415				
Phe	Asp	Thr	Gln	Val	Gly	Glu	Arg	Gly	Leu	Gln	Leu	Ser	Gly	Gly	Gln			
		420						425					430					
Lys	Gln	Arg	Ile	Ala	Ile	Ala	Arg	Ala	Met	Leu	Lys	Asn	Pro	Ala	Ile			
	435					440						445						
Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser	Ala	Leu	Asp	Ser	Glu	Ser	Glu	Lys			
	450					455				460								
Leu	Val	Gln	Glu	Ala	Leu	Asp	Arg	Phe	Met	Ile	Gly	Arg	Thr	Thr	Leu			
465					470					475					480			
Ile	Ile	Ala	His	Arg	Leu	Ser	Thr	Ile	Arg	Lys	Ala	Asp	Leu	Val	Ala			
			485					490						495				
Val	Leu	Gln	Gln	Gly	Ser	Val	Ser	Glu	Ile	Gly	Thr	His	Asp	Glu	Leu			
		500						505					510					
Phe	Ser	Lys	Gly	Glu	Asn	Gly	Val	Tyr	Ala	Lys	Leu	Ile	Lys	Met	Gln			
	515					520						525						
Glu	Ala	Ala	His	Glu	Thr	Ala	Met	Ser	Asn	Ala	Arg	Lys	Ser	Ser	Ala			
	530					535				540								
Arg	Pro	Ser	Ser	Ala	Arg	Asn	Ser	Val	Ser	Ser	Pro	Ile	Met	Thr	Arg			
545					550					555					560			
Asn	Ser	Ser	Tyr	Gly	Arg	Ser	Pro	Tyr	Ser	Arg	Arg	Leu	Ser	Asp	Phe			
			565					570						575				
Ser	Thr	Ser	Asp	Phe	Ser	Leu	Ser	Ile	Asp	Ala	Ser	Ser	Tyr	Pro	Asn			
		580						585					590					
Tyr	Arg	Asn	Glu	Lys	Leu	Ala	Phe	Lys	Asp	Gln	Ala	Asn	Ser	Phe	Trp			
	595					600						605						
Arg	Leu	Ala	Lys	Met	Asn	Ser	Pro	Glu	Trp	Lys	Tyr	Ala	Leu	Leu	Gly			
	610					615				620								
Ser	Val	Gly	Ser	Val	Ile	Cys	Gly	Ser	Leu	Ser	Ala	Phe	Phe	Ala	Tyr			
625					630					635					640			
Val	Leu	Ser	Ala	Val	Leu	Ser	Val	Tyr	Tyr	Asn	Pro	Asp	His	Glu	Tyr			
		645						650						655				
Met	Ile	Lys	Gln	Ile	Asp	Lys	Tyr	Cys	Tyr	Leu	Leu	Ile	Gly	Leu	Ser			
		660						665					670					
Ser	Ala	Ala	Leu	Val	Phe	Asn	Thr	Leu	Gln	His	Ser	Phe	Trp	Asp	Ile			
		675					680					685						
Val	Gly	Glu	Asn	Leu	Thr	Lys	Arg	Val	Arg	Glu	Lys	Met	Leu	Ser	Ala			
	690					695					700							
Val	Leu	Lys	Asn	Glu	Met	Ala	Trp	Phe	Asp	Gln	Glu	Glu	Asn	Glu	Ser			
705					710					715					720			
Ala	Arg	Ile	Ala	Ala	Arg	Leu	Ala	Leu	Asp	Ala	Asn	Asn	Val	Arg	Ser			
			725					730						735				
Ala	Ile	Gly	Asp	Arg	Ile	Ser	Val	Ile	Val	Gln	Asn	Thr	Ala	Leu	Met			
		740						745						750				
Leu	Val	Ala	Cys	Thr	Ala	Gly	Phe	Val	Leu	Gln	Trp	Arg	Leu	Ala	Leu			
	755					760						765						
Val	Leu	Val	Ala	Val	Phe	Pro	Val	Val	Val	Ala	Ala	Thr	Val	Leu	Gln			

<211> 1161
<212> PRT
<213> Arabidopsis thaliana

<300>
<308> Genbank CAA71277
<309> 1997-05-19

<400> 8

Lys	Gln	Ala	Ser	His	Arg	Val	Ala	Lys	Tyr	Ser	Leu	Asp	Phe	Val	Tyr
1				5					10					15	
Leu	Ser	Val	Ala	Ile	Leu	Phe	Ser	Ser	Trp	Leu	Glu	Val	Ala	Cys	Trp
			20					25					30		
Met	His	Thr	Gly	Glu	Arg	Gln	Ala	Lys	Met	Arg	Arg	Ala	Tyr	Leu	
		35				40					45				
Arg	Ser	Met	Leu	Ser	Gln	Asp	Ile	Ser	Leu	Phe	Asp	Thr	Glu	Ala	Ser
	50				55						60				
Thr	Gly	Glu	Val	Ile	Ser	Ala	Ile	Thr	Ser	Asp	Ile	Leu	Val	Val	Gln
65				70						75					80
Asp	Ala	Leu	Ser	Glu	Lys	Val	Gly	Asn	Phe	Leu	His	Tyr	Ile	Ser	Arg
				85					90				95		
Phe	Ile	Ala	Gly	Phe	Ala	Ile	Gly	Phe	Thr	Ser	Val	Trp	Gln	Ile	Ser
		100						105					110		
Leu	Val	Thr	Leu	Ser	Ile	Val	Pro	Leu	Ile	Ala	Leu	Ala	Gly	Gly	Ile
	115						120					125			
Tyr	Ala	Phe	Val	Ala	Ile	Gly	Leu	Ile	Ala	Arg	Val	Arg	Lys	Ser	Tyr
	130					135					140				
Ile	Lys	Ala	Gly	Glu	Ile	Ala	Glu	Glu	Val	Ile	Gly	Asn	Val	Arg	Thr
145					150					155					160
Val	Gln	Ala	Phe	Thr	Gly	Glu	Glu	Arg	Ala	Val	Arg	Leu	Tyr	Arg	Glu
			165						170					175	
Ala	Leu	Glu	Asn	Thr	Tyr	Lys	Tyr	Gly	Arg	Lys	Ala	Gly	Leu	Thr	Lys
			180					185					190		
Gly	Leu	Gly	Leu	Gly	Ser	Met	His	Cys	Val	Leu	Phe	Leu	Ser	Trp	Ala
	195					200						205			
Leu	Leu	Val	Trp	Phe	Thr	Ser	Val	Val	Val	His	Lys	Asp	Ile	Ala	Asp
	210					215					220				
Gly	Gly	Lys	Ser	Phe	Thr	Thr	Met	Leu	Asn	Val	Val	Ile	Ala	Gly	Leu
225					230					235					240
Ser	Leu	Gly	Gln	Ala	Ala	Pro	Asp	Ile	Ser	Ala	Phe	Val	Arg	Ala	Lys
			245						250					255	
Ala	Ala	Ala	Tyr	Pro	Ile	Phe	Lys	Met	Ile	Glu	Arg	Asn	Thr	Val	Thr
		260						265					270		
Lys	Thr	Ser	Ala	Lys	Ser	Gly	Arg	Lys	Leu	Gly	Lys	Val	Asp	Gly	His
	275					280						285			
Ile	Gln	Phe	Lys	Asp	Ala	Thr	Phe	Ser	Tyr	Pro	Ser	Arg	Pro	Asp	Val
	290					295					300				
Val	Ile	Phe	Asp	Arg	Leu	Asn	Leu	Ala	Ile	Pro	Ala	Gly	Lys	Ile	Val
305					310					315					320
Ala	Leu	Val	Gly	Gly	Ser	Gly	Ser	Gly	Lys	Ser	Thr	Val	Ile	Ser	Leu
			325					330						335	
Ile	Glu	Arg	Phe	Tyr	Glu	Pro	Ile	Ser	Gly	Ala	Val	Leu	Leu	Asp	Gly
	340							345					350		
Asn	Asn	Ile	Ser	Glu	Leu	Asp	Ile	Lys	Trp	Leu	Arg	Gly	Gln	Ile	Gly
	355					360						365			
Leu	Val	Asn	Gln	Glu	Pro	Ala	Leu	Phe	Ala	Thr	Thr	Ile	Arg	Glu	Asn
	370					375						380			

Ile	Leu	Tyr	Gly	Lys	Asp	Asp	Ala	Thr	Ala	Glu	Glu	Ile	Thr	Arg	Ala	385	390	395	400
Ala	Lys	Leu	Ser	Glu	Ala	Ile	Ser	Phe	Ile	Asn	Asn	Leu	Pro	Glu	Gly	405	410	415	
Phe	Glu	Thr	Gln	Val	Gly	Glu	Arg	Gly	Ile	Gln	Leu	Ser	Gly	Gly	Gln	420	425	430	
Lys	Gln	Arg	Ile	Ala	Ile	Ser	Arg	Ala	Ile	Val	Lys	Asn	Pro	Ser	Ile	435	440	445	
Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser	Ala	Leu	Asp	Ala	Glu	Ser	Glu	Lys	450	455	460	
Ser	Val	Gln	Glu	Ala	Leu	Asp	Arg	Val	Met	Val	Gly	Arg	Thr	Thr	Val	465	470	475	480
Val	Val	Ala	His	Arg	Leu	Ser	Thr	Val	Arg	Asn	Ala	Asp	Ile	Ile	Ala	485	490	495	
Val	Val	His	Glu	Gly	Lys	Ile	Val	Glu	Phe	Gly	Asn	His	Glu	Asn	Leu	500	505	510	
Ile	Ser	Asn	Pro	Asp	Gly	Ala	Tyr	Ser	Ser	Leu	Leu	Arg	Leu	Gln	Glu	515	520	525	
Thr	Ala	Ser	Leu	Gln	Arg	Asn	Pro	Ser	Leu	Asn	Arg	Thr	Leu	Ser	Arg	530	535	540	
Pro	His	Ser	Ile	Lys	Tyr	Ser	Arg	Glu	Leu	Ser	Arg	Thr	Arg	Ser	Ser	545	550	555	560
Phe	Cys	Ser	Glu	Arg	Glu	Ser	Val	Thr	Arg	Pro	Asp	Gly	Ala	Asp	Pro	565	570	575	
Ser	Lys	Lys	Val	Lys	Val	Thr	Val	Gly	Arg	Leu	Tyr	Ser	Met	Ile	Arg	580	585	590	
Pro	Asp	Trp	Met	Tyr	Gly	Val	Cys	Gly	Thr	Ile	Cys	Ala	Phe	Ile	Ala	595	600	605	
Gly	Ser	Gln	Met	Pro	Leu	Phe	Ala	Leu	Gly	Val	Ser	Gln	Ala	Leu	Val	610	615	620	
Ser	Tyr	Tyr	Ser	Gly	Trp	Asp	Glu	Thr	Gln	Lys	Glu	Ile	Lys	Lys	Ile	625	630	635	640
Ala	Ile	Leu	Phe	Cys	Cys	Ala	Ser	Val	Ile	Thr	Leu	Ile	Val	Tyr	Thr	645	650	655	
Ile	Glu	His	Ile	Cys	Phe	Gly	Thr	Met	Gly	Glu	Arg	Leu	Thr	Leu	Arg	660	665	670	
Val	Arg	Glu	Asn	Met	Phe	Arg	Ala	Ile	Leu	Lys	Asn	Glu	Ile	Gly	Trp	675	680	685	
Phe	Asp	Glu	Val	Asp	Asn	Thr	Ser	Ser	Met	Leu	Ala	Ser	Arg	Leu	Glu	690	695	700	
Ser	Asp	Ala	Thr	Leu	Leu	Lys	Thr	Ile	Val	Val	Asp	Arg	Ser	Thr	Ile	705	710	715	720
Leu	Leu	Gln	Asn	Leu	Gly	Leu	Val	Val	Thr	Ser	Phe	Ile	Ile	Ala	Phe	725	730	735	
Ile	Leu	Asn	Trp	Arg	Leu	Thr	Leu	Val	Val	Leu	Ala	Thr	Tyr	Pro	Leu	740	745	750	
Val	Ile	Ser	Gly	His	Ile	Ser	Glu	Lys	Leu	Phe	Met	Gln	Gly	Tyr	Gly	755	760	765	
Gly	Asp	Leu	Asn	Lys	Ala	Tyr	Leu	Lys	Ala	Asn	Met	Leu	Ala	Gly	Glu	770	775	780	
Ser	Val	Ser	Asn	Ile	Arg	Thr	Val	Ala	Ala	Phe	Cys	Ala	Glu	Glu	Lys	785	790	795	800
Ile	Leu	Glu	Leu	Tyr	Ser	Arg	Glu	Leu	Leu	Glu	Pro	Ser	Lys	Ser	Ser	805	810	815	
Phe	Arg	Arg	Gly	Gln	Ile	Ala	Gly	Leu	Phe	Tyr	Gly	Val	Ser	Gln	Phe	820	825	830	
Phe	Ile	Phe	Ser	Ser	Tyr	Gly	Leu	Ala	Leu	Trp	Tyr	Gly	Ser	Thr	Leu				

	835		840		845										
Met	Asp	Lys	Gly	Leu	Ala	Gly	Phe	Lys	Ser	Val	Met	Lys	Thr	Phe	Met
	850					855					860				
Val	Leu	Ile	Val	Thr	Ala	Leu	Ala	Met	Gly	Glu	Thr	Leu	Ala	Leu	Ala
	865					870					875				880
Pro	Asp	Leu	Leu	Lys	Gly	Asn	Gln	Met	Val	Ala	Ser	Val	Phe	Glu	Ile
				885					890					895	
Leu	Asp	Arg	Lys	Thr	Gln	Ile	Val	Gly	Glu	Thr	Ser	Glu	Glu	Leu	Asn
			900					905					910		
Asn	Val	Glu	Gly	Thr	Ile	Glu	Leu	Lys	Gly	Val	His	Phe	Ser	Tyr	Pro
	915						920					925			
Ser	Arg	Pro	Asp	Val	Val	Ile	Phe	Arg	Asp	Phe	Asp	Leu	Ile	Val	Arg
	930					935					940				
Ala	Gly	Lys	Ser	Met	Ala	Leu	Val	Gly	Gln	Ser	Gly	Ser	Gly	Lys	Ser
	945				950					955					960
Ser	Val	Ile	Ser	Leu	Ile	Leu	Arg	Phe	Tyr	Asp	Pro	Thr	Ala	Gly	Lys
			965					970						975	
Val	Met	Ile	Glu	Gly	Lys	Asp	Ile	Lys	Lys	Leu	Asp	Leu	Lys	Ala	Leu
		980						985					990		
Arg	Lys	His	Ile	Gly	Leu	Val	Gln	Gln	Glu	Pro	Ala	Leu	Phe	Ala	Thr
	995						1000					1005			
Thr	Ile	Tyr	Glu	Asn	Ile	Leu	Tyr	Gly	Asn	Glu	Gly	Ala	Ser	Gln	Ser
	1010					1015					1020				
Glu	Val	Val	Glu	Ser	Ala	Met	Leu	Ala	Asn	Ala	His	Ser	Phe	Ile	Thr
	1025				1030					1035					1040
Ser	Leu	Pro	Glu	Gly	Tyr	Ser	Thr	Lys	Val	Gly	Glu	Arg	Gly	Val	Gln
				1045				1050						1055	
Met	Ser	Gly	Gly	Gln	Arg	Gln	Arg	Ile	Ala	Ile	Ala	Arg	Ala	Ile	Leu
		1060					1065					1070			
Lys	Asn	Pro	Ala	Ile	Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser	Ala	Leu	Asp
	1075						1080				1085				
Val	Glu	Ser	Glu	Arg	Val	Val	Gln	Gln	Ala	Leu	Asp	Arg	Leu	Met	Ala
	1090					1095					1100				
Asn	Arg	Thr	Thr	Val	Val	Val	Ala	His	Arg	Leu	Ser	Thr	Ile	Lys	Asn
	1105				1110					1115					1120
Ala	Asp	Thr	Ile	Ser	Val	Leu	His	Gly	Gly	Lys	Ile	Val	Glu	Gln	Gly
			1125					1130						1135	
Ser	His	Arg	Lys	Leu	Val	Leu	Asn	Lys	Ser	Gly	Pro	Tyr	Phe	Lys	Leu
	1140						1145						1150		
Ile	Ser	Leu	Gln	Gln	Gln	Gln	Gln	Pro							
	1155					1160									

<210> 9

<211> 986

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic consensus sequence

<400> 9

Met	Asp	Glu	Gly	Ala	Leu	Ser	Asp	Arg	Lys	Lys	Lys	Val	Gly	Val	Leu
1				5					10					15	
Phe	Arg	Tyr	Ala	Asp	Trp	Asp	Lys	Leu	Met	Leu	Gly	Thr	Leu	Ala	Ala
			20					25					30		
Ile	Ile	His	Gly	Ser	Leu	Pro	Leu	Met	Met	Ile	Val	Phe	Gly	Glu	Met
	35						40					45			

Thr	Asp	Phe	Ala	Ser	Lys	Leu	Glu	Glu	Glu	Met	Thr	Arg	Tyr	Ala	Tyr
50						55					60				
Tyr	Tyr	Ser	Gly	Leu	Gly	Ala	Gly	Val	Leu	Val	Ala	Tyr	Ile	Gln	Val
65					70					75				80	
Ser	Trp	Leu	Ala	Ala	Gly	Arg	Gln	Ile	Arg	Lys	Ile	Arg	Lys	Phe	Phe
				85					90					95	
His	Ala	Ile	Leu	Arg	Gln	Glu	Ile	Gly	Trp	Phe	Asp	Ile	Thr	Gly	Glu
			100					105					110		
Leu	Asn	Thr	Arg	Leu	Thr	Asp	Asp	Ile	Ser	Lys	Ile	Asn	Asp	Gly	Ile
	115					120						125			
Gly	Asp	Lys	Val	Gly	Met	Phe	Phe	Gln	Val	Ala	Thr	Phe	Leu	Ala	Gly
	130				135						140				
Phe	Ile	Val	Gly	Phe	Ile	Gly	Trp	Lys	Leu	Thr	Leu	Val	Ile	Leu	Ala
145					150					155					160
Ile	Ser	Pro	Ile	Ile	Gly	Leu	Ser	Ala	Ala	Val	Trp	Ala	Lys	Ile	Leu
				165				170						175	
Ser	Phe	Ser	Lys	Glu	Leu	Ala	Tyr	Ala	Lys	Ala	Gly	Ala	Val	Ala	Glu
			180					185					190		
Glu	Leu	Gly	Ala	Ile	Arg	Thr	Val	Ile	Ala	Phe	Gly	Gly	Gln	Lys	Glu
	195					200						205			
Leu	Glu	Arg	Tyr	Gln	Lys	Leu	Glu	Ala	Lys	Lys	Ile	Gly	Ile	Lys	Lys
	210				215						220				
Ala	Ile	Ser	Ala	Ile	Ser	Met	Gly	Ala	Phe	Leu	Leu	Ile	Tyr	Ala	Ser
225					230					235					240
Tyr	Ala	Leu	Ala	Phe	Trp	Tyr	Gly	Ser	Thr	Leu	Val	Ile	Ser	Glu	Tyr
				245					250					255	
Thr	Ile	Gly	Ala	Met	Thr	Val	Phe	Phe	Ser	Ile	Leu	Ile	Gly	Ala	Phe
			260				265						270		
Ser	Val	Gly	Gln	Ala	Ala	Pro	Ile	Asp	Ala	Phe	Ala	Asn	Ala	Arg	Gly
	275					280						285			
Ala	Ala	Tyr	Ile	Phe	Lys	Ile	Ile	Asp	Asn	Pro	Ser	Ile	Asp	Ser	Phe
	290				295					300					
Ser	Gly	His	Lys	Pro	Asp	Ile	Lys	Gly	Asn	Leu	Glu	Phe	Lys	Asp	Val
305				310						315					320
His	Phe	Ser	Tyr	Pro	Ser	Arg	Glu	Val	Lys	Ile	Leu	Lys	Gly	Leu	Asn
				325					330					335	
Leu	Lys	Val	Ser	Gly	Gln	Thr	Val	Ala	Leu	Val	Gly	Ser	Gly	Cys	Gly
			340					345					350		
Lys	Ser	Thr	Thr	Val	Gln	Leu	Ile	Gln	Arg	Leu	Tyr	Asp	Pro	Glu	Gly
	355					360						365			
Val	Ile	Asp	Gly	Gln	Asp	Ile	Arg	Thr	Ile	Asn	Val	Arg	Tyr	Leu	Arg
	370				375						380				
Glu	Ile	Ile	Gly	Val	Val	Ser	Gln	Glu	Pro	Val	Leu	Phe	Ala	Thr	Thr
385					390					395					400
Ile	Ala	Glu	Asn	Ile	Tyr	Gly	Arg	Asp	Val	Thr	Met	Asp	Glu	Ile	Glu
			405					410						415	
Lys	Ala	Val	Lys	Glu	Ala	Asn	Ala	Tyr	Glu	Phe	Ile	Met	Lys	Leu	Pro
			420					425					430		
Phe	Asp	Thr	Leu	Val	Gly	Glu	Arg	Gly	Ala	Gln	Leu	Ser	Gly	Gly	Gln
	435					440						445			
Lys	Gln	Arg	Ile	Ala	Ile	Ala	Arg	Ala	Leu	Val	Arg	Asn	Pro	Lys	Ile
	450				455						460				
Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser	Ala	Leu	Asp	Thr	Glu	Ser	Glu	Ala
465				470						475					480
Val	Val	Gln	Ala	Leu	Asp	Lys	Ala	Arg	Gly	Arg	Thr	Thr	Ile	Val	Ile
				485				490						495	
Ala	His	Arg	Leu	Ser	Thr	Val	Arg	Asn	Ala	Asp	Val	Ile	Ala	Gly	Phe

1000
 900
 800
 700
 600
 500
 400
 300
 200
 100
 0

Val Lys Glu His Gly Thr His Gln Gln Leu Leu Ala Gln Lys Gly Ile
 965 970 975
 Tyr Phe Ser Met Val Val Gln Ala Gly Thr
 980 985

<210> 10
 <211> 14000
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> misc_feature
 <222> (3429)...(0)
 <223> Translation start codon

<300>
 <308> Genbank AP000386
 <309> 1999-08-03

<400> 10
 caattaatta tatcaaaatt tcgtatatca tttaaaattc atcactgatt tttgtttaga 60
 aaaaaagata gatagctatg gacatgacgt cgaatttttaa tatatcctat gtaacagtgt 120
 tcatataatc aaaaaagaaa aaataattac tattgtttgt gttctttgca acaatgcggt 180
 agaaaactcg agaccacttg ttttttctct tatgatttcg cttcagtatg atttagcaaa 240
 gcatctctac ggtacaaaaat atcttacgaa ttgacatgat ctgatccaca ctatatcaag 300
 gagttagaaa gacagaactg aaagcttctt agctcgatat gtatttcaag ctacggttca 360
 caccatctag ggagacaaga atggaagaag gcatgacgtc aagtcgaact atgcagctct 420
 cttacttaaa tgggtcgaca aacaaatacg gaaccaacta tcatccattc gagcaagcgg 480
 tgataggttg tatgactaag gtctctaaat ttggtttacg gcgaaaaatt aattgatcaa 540
 aaaatttttg gtttagtggt ttaagatctt tttcaaaaaa acgttgtatt tgatgtataa 600
 acgctatttc tttttgaaat taatttaata ttatttttgt ttgaaaagaa gaagaaacat 660
 aattcaaaaca atccctatct ttaatgttcc aacatcttaa taaatacata attaaaactc 720
 tacgatcaat actacacttc tgacgattat taaatcgcat cgtcgtacac tattattaca 780
 aagaccgttc tatctacatt aatattcttt atatattttt ttttttcatt ttttgagatt 840
 gttggagtat atatccttag attttcttgt atataaaaat agataattat aatgggataa 900
 acgaaacatc gttgaagcgc tgaagtggca ttgggtctaca ccagagcaac acaaaaaagac 960
 aaccacttca catgggtctgc tcttacattc ataaccgaat ctaagtcact ttttagttggt 1020
 tgtaattttt taactattat tccaacacct ctttttttac tttttagatt gttttataat 1080
 agaaatattt taattcctaa attaataatt aaagtaaatg taatatgagt cagtacaata 1140
 ggtgaaaaac ttaaaaagtt gacagaattt agcatttgat taaaagtgtg tgaagaagaa 1200
 gaaaaaagaa gatctttttg tatctataga tttagtgcac aacttttctc agattttcga 1260
 tatatacaag aatttaacat aagaaaaatc aagacaaatg gacctgggta taatcgggta 1320
 tctgttgtaa atattatatt tcatattctt ctccacttca taattcttat tggagttcct 1380
 tcaccaaag tttgatgttc cattaaatta atctaccact ctaataagag gtatcgtact 1440
 acaaattaca cttcataaac aagagaagaa cataaatttg aattttttta aaaaaacata 1500
 tgcgttataa caccaaacag taacggacta gctgatcctt gaatttatat tagttgcaaa 1560
 aatttatata tcgaaattga aacatgaatt tttaaaatta ttagaaaaat tatgatgttg 1620
 tctaaatgtg acattacaaa tacatgatgt tgtttaaaaa ttattataaa acaaccaaag 1680
 tttggcgtca ttctggtaaa cgcacttaat gaatttatat acagttaaag attttaacat 1740
 caaattttta aatgacaact taactaaaat ttgtatccta atatttttac tagagaaact 1800
 cacacatatt tttcaaaaca atgattagtt atatatcctt cgataatagg tattgtgtaa 1860
 aactgtgtgg tttgcacaag tgatccctcg atattctttc tgctaaagat cgacttccca 1920
 cagtttcgat atctcgggtt tgggtgcaat agcatatgct tgtttagtat gcagataatc 1980
 gtatgagaga gtcagagaga tcatctatct ttcatgatgt ggtctgagct aatcgatcat 2040
 ttgttattta tactcgataa cgttctaatt tgtatgattt ttcagccttg atctatcaca 2100
 aaatggagat aatgaaatgg tagtcataag aaaggtaatg atccccttga catgcttatt 2160
 ataacacaaa aagtaatcgc tactgattag cttaccctat gatttgaatc atataacttt 2220

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

atactaatttg	tctaagtgtg	atgatataat	gtatatgatg	tcattttataa	cttatttagta	2280
aaaaatataaa	gttcttccatc	attgtcataa	ggagtttaatt	ggatatacat	caaaaaattt	2340
cctaaattttt	tagaaattat	taatcaagtt	ataatagaaa	ttattaaata	aaatatatga	2400
tgtttttctaa	atgtgacgat	acaaaaataca	tgatgttggt	caaaacttat	tagacagaaa	2460
ccaaagtaca	tcatcattgt	gatgaagact	ttaatggatt	tatacatttc	aaagttttta	2520
acattactta	tttaaagtga	aaattgagct	aaaatttgta	tacttatatg	ttgaccatag	2580
aaatactttct	atataatttt	agacaagttg	tctatacata	tatatatata	tatatatata	2640
tatatatatg	tatttacaca	aaaatcaagt	tttcttaatg	taacaatata	aaaaatataa	2700
atatctaagt	ttgttttaaaa	attattttaga	ccaaaaaaa	aaaaaactgt	aatcaaaactt	2760
tcattggatt	atatctaact	ctaaacatga	attttttatac	aatgacatgt	tcaataaaaat	2820
tcccattttca	tagtttgacc	agaaaaaatg	cttgacagatt	tgttaaaaatc	tctacttatg	2880
caaaaccatt	gaatcacacc	agatttttgat	ttgtagtact	ttaaaaatat	taattgaaca	2940
caacataaact	ttaacaaaaa	agatctttat	aaaaaaaaa	ctttctaaaa	tgacattaat	3000
taacacaaat	tagtcaattc	aaacaatcca	ataaaatcaa	agtaaaccct	taattgtttt	3060
ttttttccat	tttacaaaa	gttaaacatt	aattacatca	actagcaaca	cgaatctctc	3120
aatctctcat	gtctcaatga	gtctttgacc	actagtcttc	atcactctct	ctccctccta	3180
tatatacca	tcacaccctt	ctccaaccac	actctcattc	tctctctctt	ctcaccactt	3240
caaaacaaga	cactcgtaaa	agagttcttg	aacttcacaa	aacaattgtc	agattttcaa	3300
gaaaaacttt	ataaaacaaa	aaacatttca	ttcttctctc	ctctctctct	cactgctcaa	3360
tgatctcggt	ttctcactaa	accaactcgt	ttcttcttac	tttctttaac	tcggatctac	3420
aaaaaaccat	gtcggaaact	aacacaaccg	atgccaaagac	tgttccagca	gaagcagaga	3480
agaagaaaga	acagagttta	ccattcttta	aactcttttc	tttcgctgat	aaatttgatt	3540
atctcttaat	gttcgttggt	tctcttggtg	ccattgttca	tggtctctcc	atgcctgtct	3600
tctttttact	ctttggtcaa	atggttaatg	gatttggtta	aaaccaaagt	gatttacatc	3660
aatgggttca	tgaagtctct	agagtaagtc	ttttttttct	ttcttggttt	tatcaaatta	3720
gatccaagat	ctgatcta	ttttgtgttt	gtgaaatttt	gcagtattct	ctatatttcg	3780
tctacttggg	tttggtcgtt	tgcttctctt	cttacgcagg	tttggttctt	acaaatccaa	3840
tcttcttttt	ctttattctt	caatcaagaa	accagtaat	taaatcacat	tataaacaag	3900
atcaagaaat	tttttgcaca	aaaaaaaaag	aaaacaagat	caagaaatta	ataaagttta	3960
gattaaacat	ttgcattatt	ttcttttttt	tggttcaacc	aattaaccca	ttaacataaa	4020
ttaaatacaga	ttattaggca	agaaatctaa	aggtgtat	tttgagt	atgtgattct	4080
tcctcaaagt	gaccattttac	tctgaaggta	aaacattttt	ttttactcta	caacaaaaac	4140
ccaacttctt	tttttgtttg	cttcaaaatt	atattaaaaa	aagaaaagctt	aaacttttgtt	4200
taaaatcttg	ttttttttgt	cattcgcttt	tgattagaac	taaaaaaacc	attttttatag	4260
aatgttggtt	acataagtga	tgtaatgggg	ttcggaacga	actttccgtg	ccaaccgctt	4320
ttatagacaa	gaaatctaaa	gggtgaattt	ttgagttttt	atgtgattct	tccacaaagt	4380
gatcattttac	tctggttaaac	atttttttac	tctacaacaa	aaaccaact	tttttatgtt	4440
tgcttcaata	aataaaaaacc	tagaaacaga	aatcacatc	tagagagaat	caaaattata	4500
tataaaaaaa	aaacacttca	aattttttaa	aatttttaata	tgtaaaaagg	ataagccaag	4560
tccacgtgat	tcatggacta	ctactttgtc	tatatcgaaa	aaaaaaaaaa	aatgggcatc	4620
tctctcacat	ttattacaac	ttcataaaaa	ttcttgtaat	aataaccata	attttttgtt	4680
aaataatttt	acagagatag	catgttggat	gtattctgga	gaaagacaag	tagcagcatt	4740
aaggaagaaa	tatcttgaag	cagtattaaa	acaagacgtt	gggttctttg	atactgatgc	4800
tagaactggg	gacattgtct	ttagtgtttc	tactgatact	cttcttgttc	aagatgccat	4860
tagtgaaaag	gtaaccattt	ttttttttac	taccaacaaa	atctagtatt	taccggcggg	4920
tcccattttt	ttcatgatca	atctttatca	aagttttggg	tccgtttatg	ttagcaaccg	4980
tcgtttatct	tgcgcctcag	gtgggttaggg	aagaaaaaag	actatagatt	gatttttttt	5040
tttgttaaaa	taaaagttgg	atttttttca	taagattgtg	aaaataagta	attttttttt	5100
acaaatttgc	tatagaaata	gaggat	ttcactat	aactaaaaact	taaatgatat	5160
agtactcatt	ttcttacatt	tttaataata	aaatagcttc	tttttttcca	ttgtgaaaaa	5220
tatttattat	gcaaaataat	gtttctttta	caagtaaaat	aaatat	tattaaaaaa	5280
aatgtgtaaa	gtaccttcaa	attttaatat	tacattttat	ctaaaactgt	tttttttttt	5340
gcaatttggg	cctgtctctg	tggtgcacgt	tgcaagcac	gagcta	gcactgggtc	5400
cctagtttc	atttatctca	tcactctttt	ttatgttga	tgaagtgaat	taataacaga	5460
gtagttgaaa	aaaaattgaa	tttctcaggt	tggaaacttt	atacattacc	tctcaacatt	5520
tttggcggga	ttagtagttg	gatttgtatc	agcatggaaa	ttagctttgt	taagtgttgc	5580
tgtgattccc	ggaatcgctt	tcgcgcggagg	tttatcgcgt	tatacactca	cgggaattac	5640

ttcaaagagc	cgtgaatctt	atgctaacgc	cggtgttata	gccgagcagg	taaaataaaga	5700
ttaaagtttg	ttccttttatt	tgtttttattt	aatgcccacc	aactgttcga	cgtaatgctt	5760
tttaactgtt	cttaggcttg	cttcttttgta	cttgatctct	caaagtattt	tttttttttt	5820
tcattgcttc	ttatgcaatc	caaagactta	aacagtgttt	tcactaattc	aaaatctttg	5880
cttttctttg	gtcgggtta	gattgatgtt	ttggtttctc	aggcaattgc	tcaagttcga	5940
actgtttatt	cttatgttgg	agagagtaag	gcacttaatg	cgtattcgga	tgcgattcag	6000
tatacgctta	agctcgggta	taaagcgggg	atggctaaag	ggttggggtt	aggatgtact	6060
tatggaatag	cttgtatgtc	atgggctttg	gtgttttggt	atgctggagt	ttttattcgg	6120
aatggacaaa	ccgatggagg	aaaggcggtt	actgctatat	tctctgctat	tgttggtgga	6180
atgtaagaat	tcaaaacata	acaatgatgt	ggttttgtgt	attttgtgtt	ttgttattac	6240
ttggatgtca	ctttttgttt	ctgtgtgtgt	ttttaggagt	ttggggcaat	ctttctcgaa	6300
tcttggggcg	tttagtaaag	gtaaagcggc	tggttataag	ttgatggaga	taattaacca	6360
gagaccgacg	ataattcaag	acccgttgga	tggaaaatgt	ttggatcaag	ttcatgggaa	6420
cattgagttt	aaagatgtga	cttttagcta	tccttcacgg	cctgatgtta	tgatcttcag	6480
gaactttaat	attttcttcc	cttctgggaa	aactgtggcg	gttggtgggtg	ggagtggctc	6540
tggaaagagt	actggtgttt	ccctcattga	gagattctat	gatccaaaca	gcggtaat	6600
gattgaattt	gttttttgct	atttgagtct	ttggctagt	actggatcat	aactttgttt	6660
attttttctt	gatgagcagg	gcaaattctg	ttggatgggt	ttgagataaa	gacgcttcag	6720
ttgaagt	tgcgtgaaca	aatcgggctt	gtgaatcaag	aacctgcgct	ctttgccact	6780
actatactag	agaacatact	ctatggaaag	cctgatgcaa	caatggttga	agttgaagct	6840
gctgcttccg	ctgcgaatgc	gcatagtttc	attacattac	ttcctaaagg	ctacgacaca	6900
cagggtataaa	tcaaatat	gattatatgt	ctagttaatg	tcattggcctt	ttgcttaatt	6960
ttttgttgaa	tgtcaataca	attagggttg	agaacgtggt	gttcaactct	cagggtggaca	7020
gaagcagaga	attgcaattg	ctaggggcat	gttgaaagac	ccaaagattc	tggtactaga	7080
tgaagctaca	agcgtctctg	atgctagctc	tgagagcatt	gttcaggaag	ctttagacag	7140
agtcatggtg	gggaggacca	ctggtgtgtg	tgctcatcgt	ctctgcacca	tcagaaatgt	7200
tgattccatt	gccgtgatac	agcaaggcca	agttgttgaa	accggaacac	atgaagaact	7260
cattgccaaa	tccggtgctt	acgcacccct	catcagggtt	caggaaatgg	ttggtactcg	7320
agatttctca	aaccgctcaa	ctcgtcgcac	tcgttcaacc	cgtttgagcc	attcactgtc	7380
aacgaaatca	ctcagtttaa	gatcaggaag	tttgaggaat	ctgagctatt	cttacagcac	7440
tggagctgat	ggtcggatag	agatgatttc	aaatgcagag	actgaccgaa	agactcgtgc	7500
ccctgaaaat	tactttctaca	ggcttctcaa	gcttaattca	ccggaatggc	cttactcaat	7560
catgggagca	gtaggctcaa	ttctttctgg	tttcattggt	cctacatttg	ctattgtgat	7620
gagcaacatg	atcgaagtct	tctactacac	agactatgat	tcaatggaaa	ggaaaacaaa	7680
agagtatgtc	ttcatctaca	ttggtgctgg	tctctatgct	gtgggtgctt	atttgatcca	7740
acattacttc	tttagcatca	tgggagaaaa	cctcacaaca	agagtaagaa	gaatgatgct	7800
ctcagggtat	tatcaaaatc	tccgtgaattt	gcttaaaatc	actttcccat	ttcttatttt	7860
ggtttcttga	ttgttcttat	ttagctatct	tgagaaacga	agttggttgg	ttcgatgagg	7920
atgaacacaa	ctcaagcctg	atcgctgcac	gttttagctac	tgatgcagca	gatgttaaat	7980
ccgctatagc	cgagagaatc	tcagtaattc	tacaaaacat	gacttcactt	ctcacatcct	8040
tcatagtcgc	cttcatagta	gaatggagag	tctcacttct	catcttaggc	acattcccac	8100
ttctagtctt	cgttaacttt	gctcaggtaa	ataatcta	ctttttactc	aaaatctttc	8160
aatattcatc	aatcattaaa	atataattgg	aatcttgc	tcaccattat	gatcttaaga	8220
aaaaacgaca	aaaggccaga	tttttataaa	tttatatttg	cttttcaaaa	gttcaaaact	8280
ttatagaaca	tggcaacaca	gccactgcct	ctacacgtgc	ttcatcttct	aactttatcc	8340
aagtttgc	ttatgtttac	aataataatc	ataaaagaat	tattaagaag	cttttttttt	8400
ctactttttg	gaaatagtgt	aagggtcaag	atcatggagc	cctcacatca	ataaatgtgc	8460
taaaaaaatt	aaaaaacagt	aggctttagt	ttactctctg	ggcatgtgtg	aagaatattt	8520
atttatatagt	ttctattgggt	actatgaccc	atagataaca	gtgttcacga	aaatagctaa	8580
gattcctctg	tcttttgctt	ctgctaaatc	tatcctactt	taagctttca	tatttacttc	8640
actctctgaa	ctctgaactg	tgatcccact	tctctcttta	tttaattctt	ttgcccataa	8700
aacctcacca	caaaaatcca	aaaaatctgc	aatttttttc	cttcttagaa	ccaatatattt	8760
atttagagtt	cttcattgggt	caaagttggt	gtctcagtgc	attatttact	tatcacagtg	8820
tgtgtgtcat	tgttttttaca	ccatccacta	gtcaatgttt	gcttgtgggt	ttctttgttt	8880
tggtagatta	ggttgtgatg	agtttttttt	ctagctgcaa	ggttcaggac	gtttcaggac	8940
tctgctttga	tatatcacca	acattttttt	acccgtgatc	taattattag	ttgaaaaaat	9000
tatcaaatag	atttcacaca	gaagaacata	gataggcatt	gtacttgtac	tgatgttgat	9060

gggatagagt	gttgcatatg	tgattttaact	atatggttct	acgtcatggt	ttagtggcga	9120
cttagacctt	tgattgtcaa	tcttattttt	tacaagtga	ctattattac	catctgttgt	9180
tctaaatcat	aagtattaat	atatgtggct	acattgcagc	aactatctct	gaaggggttt	9240
gctggagaca	cagctaaggc	tcattgcaaag	acttcaatga	ttgctgggtga	aggagtcagt	9300
aacattagaa	ccgtagcagc	tttcaatgca	cagagcaaga	ttctctcttt	gttctgtcat	9360
gagcttcgtg	tacctcagaa	aagaagctta	taccgaagtc	aaacctcggg	tttcctattt	9420
ggcctctcgc	agcttgctct	ctatggttct	gaggctttta	ttctctggta	tgggtcccac	9480
cttgtagta	aaggcgtgtc	aaccttttcc	aaagtgatca	aagtgtttgt	ggttttggtc	9540
attactgcaa	actctgttgc	tgaaactgtc	agtcttgctc	ctgaaattat	tcggggagggt	9600
gaagctgttg	gttcgggtttt	ctcgggtcttg	gacaggcaga	ccaggattga	cccggatgat	9660
gctgatgctg	atcccgtgga	gacgatccgt	ggagacattg	agttaggca	tgttgatttc	9720
gcttaccctt	caagaccoga	cgtcatgggt	ttcagggaact	ttaacctcag	aattcgagct	9780
ggacatagcc	aagctcttgt	gggcgcgagt	gggtcagggg	agagttctgt	aattgctgatg	9840
atcgagcggg	tttacgaccc	tcttgctgga	aaagtcatga	ttgatggcaa	agacatccgc	9900
cggctaaacc	tgaaatctct	aaggctcaaa	atcgggtcttg	ttcaacaaga	accagctctc	9960
ttcgcagcaa	cgatcttcga	caacatcgcc	tatggtaaag	atgggtgcaac	tgaatccgag	10020
gtaattgatg	cagctcgagc	cgcaaatgct	cacgggtttca	tcagtgggtt	acctgaaggt	10080
tacaaaactc	cagtaggcga	aagaggagtgt	cagttatcag	gtggacagaa	acagaggatc	10140
gcgatagcaa	gagctgtgct	caagaacctc	acagtgttgc	ttctagacga	agcaactagc	10200
gcactagatg	cagaatcaga	atgctgtgct	caaggagcgt	tagagagggt	catgagagggt	10260
cggaccaccg	tggtagttgc	taccgcttgc	tccaccataa	gaggtgttga	ttgcattgggt	10320
gtgattcaag	acgggaggat	tgaggagcaa	ggcagccatt	cagagctcgt	tagccgacca	10380
gagggagctt	attcaaggct	gttacagctt	caaacacata	ggatttgaag	cttgatcatg	10440
gattaaaaac	aaaaaatcgg	tttgtgtaat	tttttttata	ttaaaacttt	aatttgggaag	10500
atctctatgg	actataacga	taatatgaat	aggtgtagat	aatgaagctt	ttggagtgtt	10560
tatgaagggt	ctttaattaa	gggttatttt	tttcgcattt	tgcttatgtg	cccgttttgg	10620
aacataagac	tgaactatgt	tttcggtatg	tttttaattt	atgcctcgaa	acaaaacaaa	10680
atcactctaa	tacttttggt	gatcaaaatt	tccctaaaac	atctgggttca	tgactagaat	10740
tgatcggaca	tgttctttta	gggggtttgat	ctcgcattgt	ctaatacagg	tatgtggtct	10800
tacagcgact	taatcaattt	agaccggcta	tcaaaattag	attcagccta	aacagcaaaa	10860
ttgtcttata	taatataccta	aatacacagac	cggatttttt	ttttttataa	actctatcgg	10920
ctgatccggg	caaccccggg	agaaacgcac	tgaagtgttt	catagtggac	ttcaaaccca	10980
aattttgcag	aaattgagat	gagaatgagc	tctacctcgc	attgacatgt	tatttcgttc	11040
ttgagatctt	aagtaatttt	ttgagatctt	aagtaatttt	ttgggtattca	agttccttat	11100
ttttttgogg	ttactcgtct	tgacatgtta	tttcgttctt	gtgctactta	cataccacaa	11160
tatatgtagg	tttctaaaca	tatataatag	gaatgttgtt	ctatataatg	gttgtttgat	11220
acgtaaagag	taattctgggt	ctacgtagag	tctatctctt	gtttcttggt	gagattttga	11280
ctggactgca	tggtaatcct	cgcgcattgt	gtttagtgtt	cgggtgcttc	attcatttgg	11340
ccagcgtcaa	atcttctaga	agactctctg	gtcttttggt	caaaggcctc	cgatttttcc	11400
atgacaagtt	cattcatatt	tttttgtgac	aacttgattg	ttgttcattt	ccgaatatcc	11460
catgaacgta	tcaaaactca	taaaatcgat	tgttgtcact	ccctttaaga	tgagatcctc	11520
aacggcagct	tcaccactag	ccatgttttc	acacatgac	aactatctca	catttttact	11580
aaggctttta	gcaggaatat	tacttctaga	attttaacaa	tctagaattt	ctagaggaaa	11640
tgttcacatg	tatatatttt	ggtatatgat	ctagagaact	tatataagat	tatcaatgtt	11700
taatatactc	aacaaatggg	ttcttctatt	ttatcttcta	attgttgggg	ccttgaatta	11760
tttagttctt	tttactattt	ccttttcttt	tatttggtgt	gtggagccat	aaagttccaa	11820
taatttctat	gaaaagaaat	gaataaaatc	taaagaaacc	aactgtgaag	ctctttaatc	11880
ttattctatc	tgttggtatc	gtttgccttt	atctgctttg	catgtggttt	tgctgtttta	11940
gataaaggcc	atatcttctc	ctaattcttt	ttccaaatca	acacatcaaa	tatataaaaag	12000
actccaaata	ttttgaaatg	aatcataatt	tattttaatta	caaaaaaact	taaagaattc	12060
aagacctgcg	atcacaagac	aaatacatgt	aaggaattta	atattttaaga	taccaaccgc	12120
taataaataa	gactcacaat	ttctaaaaat	taaaacgaaa	aaaagaaaaa	attagtcatc	12180
attgcatatc	ttagtgaaca	aaaatcaata	agaaaaggat	aatatacttt	ttaccaacat	12240
cacataaacc	taaaagcatc	tcacacaaaa	agtaaattta	aaactcagaa	aaagctgaat	12300
acaagttagt	aaatttatgg	aaagtgttag	tgtgttagtg	ttgttatcat	tacaacttta	12360
aaataattat	gaacatgaaa	aaatagattg	aacatgaagc	agcattggga	aggactcgaa	12420
tgtccctact	catactgaca	aaacctcaag	tgttccagtg	cgttaggttt	cactttccag	12480

tagtatgtgt	atctgctttt	tttgtatata	ttaatgtaat	agatagtgtg	gagaatcatg	12540
aggaacattt	tgtaatcata	gaactaaggc	ataattttga	tcgtgtgaaa	taatactatt	12600
atttttggta	acaaggatga	tctggaccga	attccgacta	attttccagg	tgaatatagt	12660
catcaagtaa	cttgatcacc	ggatttttaa	aatgaattat	gaacatgata	tcattttcaac	12720
taatcaaata	aaaatgaaaa	aattcatgtg	ggaatcatat	ccctcaaatc	tgttaccacc	12780
aaactacttt	catttggttt	gtgtgagaaa	tataagaatt	atcaaacaaa	gaaaaagggt	12840
ttgcttaatt	ataatgatat	gaggttcatc	aaatttatat	accactaaaa	tattcctaca	12900
ctcatccaaa	tattattctc	gcttatgcag	ttttaatcta	atgtacaaat	catatccata	12960
agcgcattgg	ccgatgctaa	aattttggca	ttagttgttg	atttgttttt	cttgagtctt	13020
atgttcaaaa	aatatatttg	ataatttaat	gtataaatta	taccataaat	cacattgatc	13080
aatgcttaaa	gttttgtatt	agttgttgat	tttttgtttg	tttgctgtgt	ttgaattttc	13140
ataaatgaaa	agtaaatacg	ataatcatat	ttcaaaacgt	aaagggtgta	attagtagtc	13200
taaatggtaa	aatataaaca	gtatattata	aaaattacta	aaatggtttt	gtaaaaaaaa	13260
tatgctattt	gtattataat	gaaattcaaa	aatttttaat	agaacgtata	atctctgcac	13320
aaagagggtt	tgagggtgta	taattcatga	agtaaatttt	atttactgac	gggtaagttg	13380
tgaaaaagtt	tgaagactat	tttttttgct	tttcacagag	aaaactactt	tcctttttta	13440
ttgtatgatg	agaaggcaaa	agtgcagaca	tgtgctttct	tttctcccat	tttcaacaat	13500
gtcactcggt	gtattattca	tatttttagca	aactggttat	atctatatct	atcaatcatt	13560
tcagaacatc	atatccatca	gtttttggac	attgctacat	acgttagtat	tgatgtacca	13620
gttaccctaa	caggcctttt	catagtgtgg	cagaacacgt	gagggtgtgat	atatgcggat	13680
gaattctatg	ttctgcattt	tgttaccatt	catataaaaag	tattgtttta	gttgtgctgc	13740
ggtttaagtc	ctaatacaga	tttatttgta	gtgttatggt	tataaaccac	attggaaccc	13800
aaactctaag	agaatataag	atcaatgtgt	aattaataaa	ttttatagtc	attcggataa	13860
aacttagaaa	gataacaaaa	gtaagaacga	gtatttttta	gcgaataact	tttagatatt	13920
ctttgtattg	atcattattg	acaaaccatt	aaacattttt	gtagacccta	aaatccgata	13980
tccaaaaaac	aattcttatt					14000

<210> 11

<211> 29

<212> DNA

<213> Arabidopsis thaliana

<400> 11

cactgctcaa tgatctcggt ttctcacta

29

<210> 12

<211> 30

<212> DNA

<213> Arabidopsis thaliana

<400> 12

cttgaatcac accaatgcaa tcaacacctc

30

<210> 13

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 13

cattttataa taacgctgcg gacatctac

29

<210> 14

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 14

tttctccata ttgaccatca tactcattg

29

tttctccata ttgaccatca tactcattg